9. НЕЧЕТКИЕ МНОЖЕСТВА В СИСТЕМАХ ОСНОВАННЫХ НА ЗНАНИЯХ

9.1. Основные понятия и определения

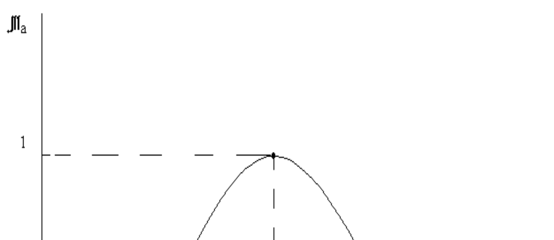
При работе с экономическими и статистическими данными как в системах принятия решений, так и в базах данных и знаний, возникают ситуации когда мы не можем точно количественно описать какие-либо объекты или явления. Например, на вопрос «Каков спрос в Санкт-Петербурге на мебель для кухни отечественных производителей?» можно получить лишь качественный ответ - «высокий», «низкий», «ниже, чем на итальянские кухни» т.п. Даже если нам удастся добыть цифру 254 кухни в месяц, то, очевидно, доверие у нас будет довольно низкое, так как это значение может не учитывать ряд мелких производителей, продавцов, сезонность и т.д.

Поэтому для успешного принятия решения мы должны научиться работать с качественными данными и нечетко определенными понятиями. В связи с этим возникают вопросы:

- как сложить два качественных описания или найти их среднее (например, спрос на отечественные кухни в Москве и Санкт-Петербурге);
- как в базе данных найти все города, где спрос на кухни между «низким» и «очень низким», т.е. как выполнить фильтрацию при таких данных.

Для ответа на подобные вопросы используется специальный математический аппарат называемый *теорией нечетких множеств*, который, как показали два последних десятилетия, оказался достаточно жизненным, широко и успешно применяемым в самых различных областях. Рассмотрим основные определения и положения теории нечетких множеств.

<u>Нечетким множеством Ã</u> на множестве U называется совокупность пар $\tilde{A} = <{}_{A}(u), u>$ где ${}_{A}(u)$ - функция принадлежности нечеткого множества Ã, а u - носитель нечеткого множества Ã.



Рассмотрим пример. Пусть требуется определить нечеткое понятие «Студент учится хорошо». Обозначим это понятие Ã. При определении этого понятия совершенно очевидно, что если у студента средний балл 4, то его можно назвать хорошо успевающим. Однако если средний балл 4.4 , то студент учится тоже «хорошо», но уже с некоторым сомнением - а может «отлично». При среднем балле 3.7, его тоже можно отнести к категории учащихся на «хорошо», но с еще большим сомнением - а может «удовлетворительно». Наше представление (знание) о понятии «Студент учится хорошо» можно представить графически (рис. 9.1).

График показывает наше сомнение или уверенность в том, что студента с некоторым средним баллом можно отнести к “хорошо успевающий”. Здесь средний балл <u> является *носителем нечеткого множества*.

А степень уверенности в том, как значение <u> принадлежит понятию «хорошо» и есть *функция принадлежности*. Таким образом, нечеткое множество Ã есть не что иное, как попытка количественно описать качественное выражение «хорошо».

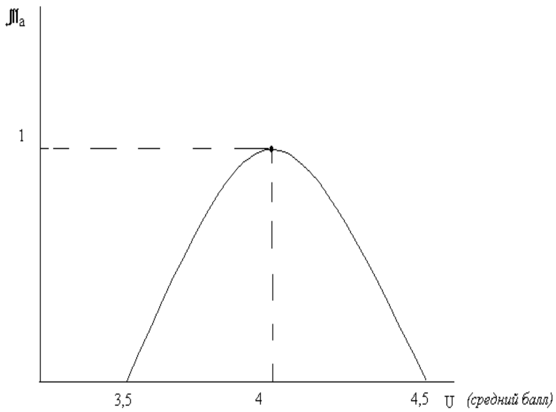
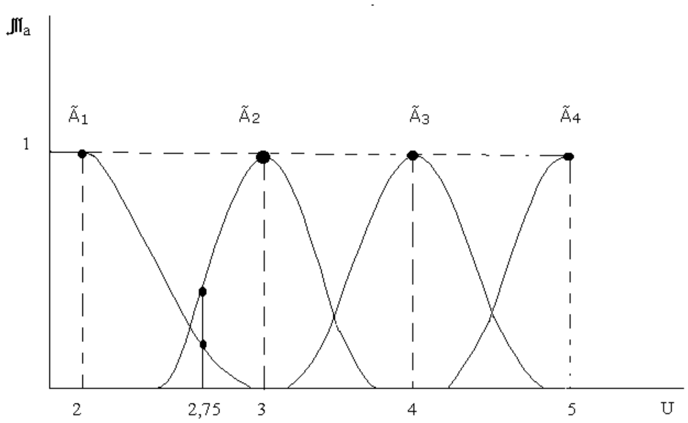


Рис. 9.1 Функция принадлежности

Почему функция принадлежности $_{A}(u)$ [0,1]? Это связано с тем, что нам удобно выражать свое отношение в процентах. Например, на 90% доверяю. А если есть абсолютная уверенность, то это 100% доверие или 1. Поэтому и область изменения значений функции принадлежности от 0 до 100% или [0, 1].

Рассмотрим теперь понятие <u>лингвистическая переменная</u>. Это переменная, которая может принимать нечеткие значения. Поясним это понятие на примере. Определим лингвистическую переменную X, характеризующую успеваемость студентов. Эта переменная может принимать одно из четырех возможных нечетких значения

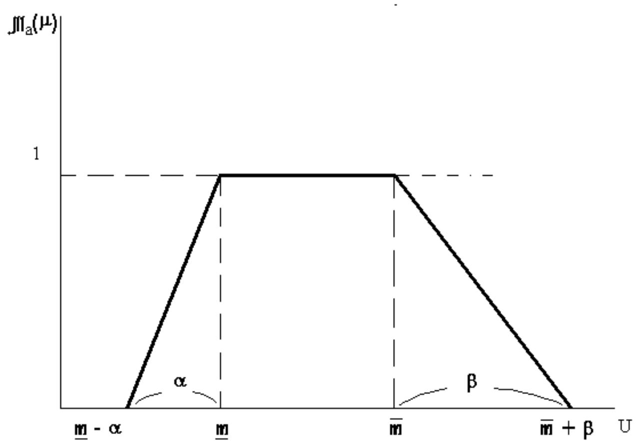
$$X = \{\tilde{A}_1, \tilde{A}_2, \tilde{A}_3, \tilde{A}_4\}$$



где $\tilde{A}_1$ = «неудовлетворительно»; $\tilde{A}_2$ = «удовлетворительно»; $\tilde{A}_3$ = «хорошо»; $\tilde{A}_4$ = «отлично». Для того, чтобы работать с лингвистической переменной X мы должны количественно определить все нечеткие множества $\tilde{A}_i$ (рис. 9.2).

Из этого рисунка видно, что графики функций принадлежности пересекаются, что говорит о том, что мы затрудняемся сказать точно, к какому нечеткому значению отнести определенный средний балл. В таких случаях обычно используют максимальное значение $_{Aiu}$, например, средний балл 2,75 более подходит к значению «удовлетворительно», чем «неудовлетворительно».

Рис. 9.2. Нечеткие множества $\tilde{A}_i$



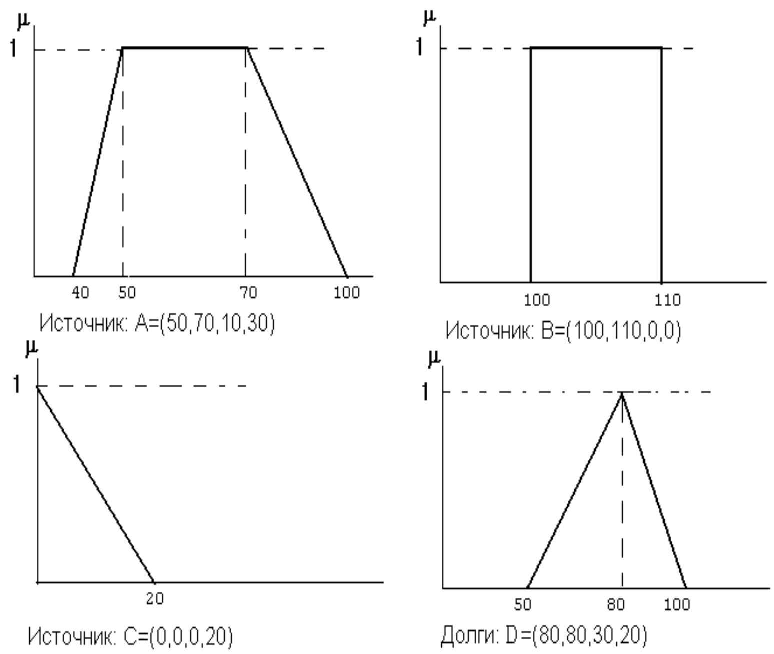
Следует отметить, что в большинстве случаев функции принадлежности строятся субъективно по результатам опроса экспертов, поэтому они являются в некотором смысле «приближенными», т.е. не абсолютно адекватно отражающими явление или объект. Собственно говоря, из субъективности следует, что абсолютной адекватности в принципе и не существует. Поэтому, очевидно, нужно выбирать такую функцию, с которой можно было бы как можно проще вести расчеты. Такими функциями являют-ся *трапециевидные* (рис. 9.3) функции. Тогда $_A(u)$ харак-теризуется четверкой ($\underline{m}$, m, ,). Как частный случай при $\underline{m}$ = m имеем треу-гольную форму функции.

Рис. 9.3. Трапециевидная функция принадлежности

9.2. Арифметические операции над нечеткими переменными

Прежде чем определить простейшие арифметические операции над нечеткими переменными рассмотрим пример. Пусть в рамках составления проекта бюджета рассматриваются различные источники финансирования. Причем некоторые из них характеризуются неточностью оценки денежных сумм на день оценивания, а другие малой надежностью. Кроме того, из бюджета необходимо отдать долги, количество которых также неточно, так как зависит от того, потребует ли кредитор все или только часть в следующем финансовом периоде.

- *Источник A*: финансирование обеспечивается, его сумма может изменяться от 40 до 100 млн. в зависимости от конъюнктуры, но с наибольшей вероятностью можно ожидать поступления в сумме от 50 до 70 млн.
- *Источник B*: источник надежен и разумно полагать, что финансирование будет предоставлено и составит сумму 100 - 110 млн.
- *Источник C*: источник ненадежен, а если и даст, то не более 20 млн.
- *Долг D*: плата за кредиты 50 - 100 млн., но наиболее вероятна выплата 80 млн.



Таким образом, имеем три источника поступлений и один источник расхода. Построим на основе их описаний трапециевидные функции принадлежности для каждой из четырех нечетких переменных (рис. 9.4).

После задания всех нечетких переменных, встает задача определения суммы всего бюджета, которая также будет нечеткой величиной. А для этого надо уметь выполнять простейшие арифметические операции над нечеткими переменными.

Рис. 9.4. Проект бюджета

Определение этих операции рассмотрим для случая двух нечетких переменных $\tilde{A}_1$ и $\tilde{A}_2$, которые заданны своими трапециевидными функциями принадлежности вида

$$\tilde{A}_1 = (\underline{m}_1,\quad m_1,\ \ _1,\ \ _1),$$
$$\tilde{A}_2 = (\underline{m}_2,\quad m_2,\ \ _2,\ \ _2).$$

Результатом операции будет также нечеткая переменная $A = (\underline{m}, m, ,)$, которая также имеет трапециевидную функцию принадлежности, параметры которой определяются в зависимости от вида арифметической операции (табл.9.1)

Таблица 9.1

Тип операции	Зависимости параметров функций принадлежности
$A = \tilde{A}_1 (+) \tilde{A}_2$	$\underline{m} = \underline{m}_1 + \underline{m}_2,\quad m = m_1 + m_2,$ $= _1 + _2,\quad = _1 + _2.$
$A = \tilde{A}_1 (-) \tilde{A}_2$	$\underline{m} = \underline{m}_1 - m_2,\quad m = m_1 - \underline{m}_2,$ $= _1 + _2,\quad = _1 + _2.$
$A = \tilde{A}_1 (x) \tilde{A}_2$	$\underline{m} = \underline{m}_1 * \underline{m}_2,\quad m = m_1 * m_2,$ $= \underline{m}_1 * \underline{m}_2 - (\underline{m}_1 - _1)(\underline{m}_2 - _2),$ $= (m_1 + _1)*(m_2 + _2) - m_1 * m_2$
$A = \tilde{A}_1 (/) \tilde{A}_2$	$\underline{m} = \underline{m}_1 / \ m_2,\quad m = m_1 / \underline{m}_2,$ $= (\underline{m}_1 * _2 + m_2 * _1) / (m_2^2 + m_2 * _2)$ $= (\underline{m}_2 * _1 + m_1 * _2) / (\underline{m}_2^2 - \underline{m}_2 * _2)$

На основе приведенных выше описаний арифметических операций можно для рассматриваемого примера определить оценку бюджета без учета долгов (Ф) как сумму трех источников финансирования. Причем результат будет также нечеткой переменной

$$Ф = A\ (+)\ B\ (+)\ C = (50+100+0,\quad 70+110+0,\quad 10+0+0,\quad 30+0+20) = (150,\ 180,\ 10,\ 50)$$

с трапециевидной функцией принадлежности, приведенной на рис. 9.5. Для получения полной оценки предполагаемого бюджета необходимо из полученного результата вычесть предполагаемые платы по кредитам. При этом бюджет с учетом долгов () также будет являться нечеткой переменной (рис. 9.6):

$$= Ф\ (-)\ D = (\ 150 - 80,\ 180 - 80,\ 10 + 20,\ 50 + 30\ ) = (70,\ 100,\ 30,\ 80),$$

функция принадлежности которой также имеет трапециевидный вид и приведена на рис.9.6.

Рис. 9.5. Бюджет без учета долгов *Рис. 9.6. Бюджет с учетом долгов*

Таким образом, в бюджете может быть сумма от 40 до 180млн., но с наибольшей степенью уверенности можно говорить о суммах от 70 до 100млн.

9.3. Операции нечеткой фильтрации и выбора

Определение этих операций также начнем с рассмотрения примера. Пусть создается база данных по бюджетам различных фирм или подразделений. При этом любой из бюджетов может быть оценен как «малый» (A_1), «средний» (A_2) или «большой» (A_3). Т.е. мы имеем дело с лингвистической переменной «объем бюджета» ($\tilde{A}$), которая может принимать одно из трех нечетких значений (A_1, A_2, A_3).

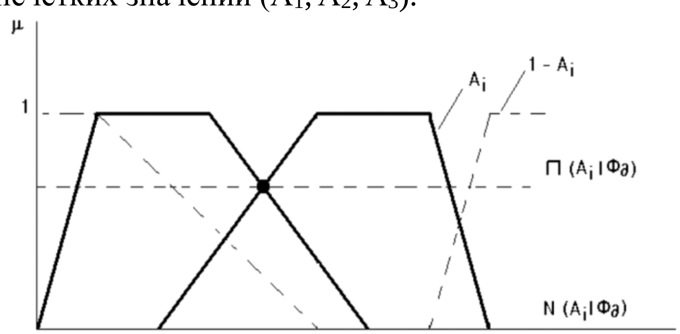


Рис. 9.7. Графическая интерпретация понятий возможности и необходимости

Предположим, что функция принадлежности для каждой из переменных были оценены следующим образом

A_1 = «малые» = (0, 50, 0, 50);
A_2 = «средние» = (80, 150, 20, 20);
A_3 = «большие» = (200, 250, 20, 20).

Теперь встает задача - к какому из классов бюджетов можно отнести значение бюджета, полученного в предыдущем примере. Т.е. необходимо определить является ли наш бюджет «малым», «средним» или «большим».

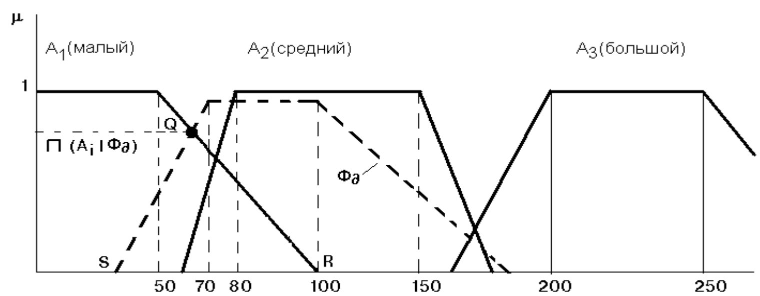
Эта задача относится к классу задач <u>нечеткой фильтрации</u> в базах данных или -отбора. Для ее решения вводится два показателя:

- $\Pi\ (A_i Ф) = \sup \min\ ({}_{ф}(u),\ {}_{Ai}(u))$ - это *возможность*, что нечеткое множество Ф принадлежит значению A_i атрибута $\tilde{A}$.
- $N\ (A_i Ф) = \inf \max\ (1 - {}_{ф}(u),\ {}_{Ai}(u))$ - это *необходимость*, что нечеткое множество Ф принадлежит значению A_i атрибута $\tilde{A}$.

Для всех значений атрибута Ã мы находим пары (П, N), а затем по максимальному значению этой пары находим принадлежность нечеткого множества к тому или иному значению атрибута. На рис. 9.7 показана суть этих показателей. Рассмотрим подробнее эти два показателя.

Показатель $П(A_i Ф)$ характеризует *возможность совпадения* (мягкое решение или мягкий вывод – отсюда английское soft computing). Фактически, любую функцию принадлежности можно рассматривать как распределение возможностей. Например, из рис. 9.6 видно, что возможность получения суммы 120 млн в бюджете Ф равна 0.75. В отличие от данной возможности, возможность $П(A_i Ф)$ является условной и вычисляется на основе принципа обобщения Заде (приведенные выше формулы).

Пусть $_{A1}(u)$, $_{A2}(u)$, и $_{A3}(u)$ имеют вид, приведенный на рис. 9.8. На этом же рисунке изображена функция принадлежности для Ф. Рассмотрим геометрическую интерпретацию определения $П(A_1 Ф)$:



- min ($_{ф}(u)$, $_{A1}(u)$) – представляет собой треугольник SQR, т.е. это нечеткое множество с функцией принадлежности ограниченной точками SQR. Мы для каждой точки на оси взяли наименьшее значение из двух $_{ф}(u)$ и $_{A1}(u)$.
- sup min ($_{ф}(u)$, $_{A1}(u)$) – это точка Q, т.е. наибольшее значение M_{SQR} во всех точках U=[0,). Это примерно 0.75, т.е. $П(A_1 Ф) = 0.75$.

Рис. 9.8. Геометрическое определение П(A₁Ф)

Значение $П(A_1 Ф)$ может быть вычислено и аналитически на основе приведенных выше формул принципа обобщения Заде. Пусть

$$Ф = (\underline{m}_1, m_1, {}_1, {}_1),$$
$$A_1 = (\underline{m}_2, m_2, {}_2, {}_2).$$

Тогда

$$П(A_1 Ф) = \min \{\max (0, \min (1, 1 + (m_1 - \underline{m}_2) / ({}_1 + {}_2))), \max (0, \min (1, 1 + (m_2 - \underline{m}_1) / ({}_2 + {}_1)))\}.$$

Для рассматриваемого примера имеем:

$$Ф =(70, 100, 30, 80);$$
$$A_1 = (0, 50, 0, 80).$$

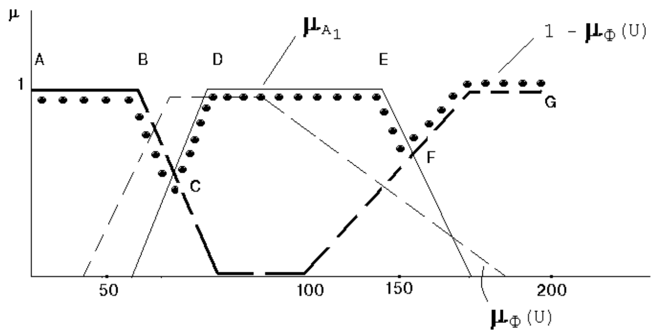
Используя вышеприведенное выражение, получим

$$П(A_1 Ф) = \min\{ \max (0, \min(1, 1 + (100- 0) / (80 + 0))), \max (0, \min(1, 1 + (50 -70) / (30 + 50)))\} =$$
$$= \min\{ \max (0, \min(1, 1 + 100 / 80)), \max (0, \min(1, 1 - 20 / 80))\} =$$
$$= \min\{ \max (0, 1), \max (0, 1 - 0,25)\} = \min\{ 1 , 0,75 \} = 0,75 ,$$

что соответствует графическому расчету. Аналогично можно найти $П(A_1 Ф)$ для A_2 и A_3. Даже из рисунка видно, что П(A_2Ф) = 1; П(A_3Ф)   0.2

Показатель N(A_iФ) характеризует *необходимость совпадения*, т.е. значение бюджета A_i обязательно принадлежат Ф (жестокое принятие решения, даже сверхжесткое). Эта величина используется в двух случаях:

- когда П(A_i Ф) = П(A_jФ), i j, т.е. возможности одинаковы и, следственно, неразличимы;
- при более жестком отборе.



Рассмотрим суть этого показателя на нашем примере при вычислении N(A_2Ф). Как и ранее начнем с геометрической интерпретации определения П(A_1Ф):

- Сначала найдем 1 - $_{\phi}$(u)
- Затем   max (1 - $_{\phi}$(u), $_{A2}$(u)). Это ломаная ABCDEFG (на рис. 9.9. она обозначена точками).
- Вычислим inf max (1 - $_{\phi}$(u), $_{Ai}$(u)). Это будет нижняя граница ломаной ABCDEFG. Очевидно, что это точка C.
- Тогда N(A_2 Ф) будет соответствовать точке. Это примерно 0.20.

Выполняя аналогичные действия, из рис. 9.9 можно получить, что N(A_1Ф)= 0   и   N(A_3Ф) = 0

Рис. 9.9. Геометрическое определение N(A_1Ф)

Используя формулы, вытекающие из принципа обобщения Заде можно провести аналитический расчет показателя необходимости. Если принять, что A_2 = ($\underline{m}_2$, m_2, $_2$, $_2$), тогда

$$N(A_2Ф) = \min \{\max (0, \min (1, (\underline{m}_1 - \underline{m}_2 + {}_2) / ({}_1 + {}_2))),$$
$$\max (0, \min (1, (m_2 - m_1 + {}_2) / ({}_1 + {}_2)))\}.$$

Для данного примера:

Ф = (70, 100, 30, 80).

A_2 = (80, 150, 20, 20).

Тогда

N(A_2Ф) = min {max (0, min (1, (70-80+20)/(30+20)),

max (0, min (1,(150-100+20)/(100)) =

= min {max (0, min (1, 0,2), max (0, min (1, 0.7) = 0.2,

что соответствует графическим преобразованиям, полученным из рис. 9.9. Таким образом, после проведения всех вычислений мы имеем три значений показателей возможности и необходимости (П, N) для каждого из значений A_1, A_2, A_3 лингвистической переменной «объем бюджета» (Ã):

для A_1="малый"   (0,75; 0,0)

для A_2="средний" (1,00; 0,2)
для A_3="большой" (0,20; 0,0)

из которых очевидно, что рассчитанный нами бюджет необходимо отнести к «среднему» с возможностью 1 и необходимостью 0.2