МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

Механико-математический факультет

# Модели поиска информации в нечеткой среде

**А.П. Рыжов**

Москва 2004 год

А.П. Рыжов

**Модели поиска информации в нечеткой среде**

В работе рассматриваются вопросы поиска информации в нечеткой среде. Под средой поиска информации понимается пара запрос, база данных. Рассматриваются варианты, когда запрос и/или база данных могут быть четкими или нечеткими. Приводятся и исследуются математические модели описания человеком объектов предметной области некоторой базы данных и поиска информации в нечетких (лингвистических) базах данных. На основе полученных результатов формулируются методики выбора оптимального множества значений качественных признаков для двух критериев:

- минимизация трудностей описания человеком реальных объектов;
- минимизация потерь информации и информационных шумов при поиске информации в нечетких (лингвистических) базах данных.

Работа ориентирована на студентов, аспирантов и специалистов, занимающихся вопросами обработки информации в рамках человеко-компьютерных систем.

Рецензент — профессор В. Б. Кудрявцев

# Оглавление

**Глава 1. О поиске информации в нечеткой среде** **5**
1.1 Существующее положение . . . . . . . . . . . . . . 5
1.2 Классификация задач поиска информации в нечеткой среде . . . . . . . . . . . . . . . . . . . . . 8
**Глава 2. Семантические пространства** **10**
2.1 Понятие лингвистической переменной . . . . . . . 10
2.2 Концепция полного ортогонального семантического пространства (ПОСП) . . . . . . . . . . . . 12
2.3 Степень нечеткости ПОСП . . . . . . . . . . . . . . 15
2.4 Некоторые свойства степени нечеткости ПОСП . . 18
**Глава 3. Нечеткий запрос — нечеткая БД** **22**
3.1 Лингвистические базы данных . . . . . . . . . . . 22
3.2 Потери информации и шумы . . . . . . . . . . . . 24
3.3 Методика выбора оптимального множества значений качественных признаков . . . . . . . . . . . 30
3.3.1 Устойчивость методики выбора оптимального множества значений . . . . . . . . . . 41
**Глава 4. Нечеткий запрос — четкая БД** **61**
4.1 Описание нечеткого лингвистического интерфейса 63
4.2 Блок формализации понятий пользователя . . . . 63
4.3 Блок поиска информации . . . . . . . . . . . . . . 64
4.4 Блок анализа результатов поиска . . . . . . . . . . 65
**Прил. A. Доказательства теорем** **67**
A.1 Доказательство леммы 1 . . . . . . . . . . . . . . . 67
A.2 Доказательство теоремы 1 . . . . . . . . . . . . . . 67
A.3 Доказательство теоремы 2 . . . . . . . . . . . . . . 69
A.4 Доказательство теоремы 4 . . . . . . . . . . . . . . 71
**Прил. B. Функции принадлежности** **73**
B.1 Классификация методов построения функций принадлежности . . . . . . . . . . . . . . . . . . . . . 73
B.2 Прямые методы для одного эксперта . . . . . . . . 75
B.3 Косвенные методы для одного эксперта . . . . . . 77
B.4 Прямые методы для группы экспертов . . . . . . . 79
B.5 Косвенные методы для группы экспертов . . . . . 81
B.6 Параметрическое задание функций принадлежности . . . . . . . . . . . . . . . . . . . . . . . . . . 83

Данная работа подготовлена в рамках проекта № 431 (направление 2.1) Федеральной целевой программы "Интеграция Создание совместного научно-учебного центра "Интеллектуальные системы и нечеткие технологии. Основу работы составил курс лекций, читаемых автором в течение ряда лет для студентов и аспирантов механико-математического факультета и факультета вычислительной математики и кибернетики МГУ им. М.В. Ломоносова.

В работе используются некоторые свойства модели описания человеком объектов предметной области, представленные в монографии автора "Элементы теории нечетких множеств и измерения нечеткости", Москва, Диалог-МГУ, 1998 г., выполненной в рамках этого же проекта. Формулировки соответствующих теорем приведены в разделе 2. Однако, учитывая практическую недоступность данной книги, вышедшей ограниченным тиражом и раскупленной в течение нескольких месяцев, доказательства основных теорем приведены в приложении А. Обзор методов построения функций принадлежности, важных для создания приложений, но не являющихся центральным для данной работы, приведен в приложении В.

Автор хотел бы выразить свою искреннюю благодарность научному руководителю проекта заведующему кафедрой математической теории интеллектуальных систем механико-математического факультета МГУ им. М.В. Ломоносова академику профессору В.Б. Кудрявцеву, который был инициатором этой работы, за его труд по неоднократному ее прочтению, редактированию и весьма ценные замечания.

Концепция работы и ее содержание неоднократно обсуждались с директором Центра компьютерных технологий в образовании РГГУ профессором А.С. Строгаловым и заместителем заведующего отделом искусственного интеллекта ВЦ РАН членом-корреспондентом МАИ А.Н. Аверкиным. Автор хотел бы также их поблагодарить за проделанную большую работу.

Выпуск работы был бы не возможен без финансовой поддержки Федеральной целевой программы "Интеграция".

*А.П. Рыжов. Москва, 2004 год.*

**Глава 1.**

# О поиске информации в нечеткой среде

## 1.1. Существующее положение

Применение аппарата теории нечетких систем в задачах поиска информации переживает в настоящее время период бурного роста.

Вопросы математически корректного описания плохоформализуемой, слабо структурированной информации и ее обработки всегда находились в поле зрения специалистов по нечетким системам. Более того, теория нечетких множеств по крайней мере на начальном этапе своего развития и виделась как средство описания такой информации. Не случайно первой "нечеткой" теорией была теория лингвистической переменной Заде [10]. Естественным вопросом в рамках обработки нечеткой информации явился вопрос поиска информации в базах данных, описания объектов которых есть лингвистические описания объектов предметной области. Систематизированное изложение результатов в этой области впервые было предпринято в монографии [38]. После этого вопрос нечетких баз данных с теоретической точки зрения считался "закрытым", так как обобщения теории баз данных на случай нечетких данных были получены, и интерес специалистов в области нечетких систем к нему пропал. Однако, как это часто бывает, данная работа осталась не замеченной специалистами в области баз данных.

Развитие технологий обработки информации и экспансия их в различные предметные области привело к выделению класса систем, для которых источником информации является человек. Самым известным типом такого рода систем являются экспертные системы. Попытки использовать для проектирования такого класса систем и разработки алгоритмов обработки такой информации, наработанные в других областях формально-логические методы оказались не эффективными. Широким фронтом проводились работы по разработке методов обработки неопределенной информации. При этом основное внимание уделялось логическому выводу, объяснению вырабатываемых рекомендаций и т.п. Проблемы собственно поиска информации в нечетких базах данных долгое время оставались на периферии проводимых исследований. И только в начале 90-х годов они вновь привлекли значительный интерес как теоретиков, так и прикладников. Так на конференции Североамериканской ассоциации нечетких систем была одна секция по нечетким методам поиска информации, на которой было представлено 3 доклада [47], на Всемирном конгрессе Международной ассоциации нечетких систем в Сеуле была также одна секция по нечетким методам поиска информации, на которой было представлено 5 докладов [51], на первом Европейском конгрессе по нечетким и интеллектуальным технологиям таких докладов было 17 на 3 секциях [50]. Стали проводиться специализированные конференции по данной проблеме - International Symposium on Intelligent Data Analysis (IDA-95) Baden-Baden, Germany 17th-19th August 1995; International Workshop on Flexible Query-Answering Systems FQAS'96: Roskilde University, Denmark May 22-24, 1996. Появился Международный журнал Intelligent Data Analysis, посвященный данной проблематике.

Причин сложившейся ситуации несколько. Их можно разделить на "внутренние" и "внешние". К первым можно отнести накопившийся к настоящему времени опыт (как позитивный, так и негативный) разработки и использования таких систем. Именно в рамках такого опыта могли возникать вопросы типа: "Можно ли предложить такое правило описания реальных объектов, чтобы человек - источник информации описывал объекты с минимальными трудностями?", "Как формировать описания объектов, чтобы разные источники описывали их более или менее одинаково?", "Как описывать объекты, чтобы обеспечить максимально хорошие показатели качества поиска информации?" и т.п.

К "внешним" можно отнести следующие причины. Во-первых, опыт практического использования экспертных систем и других "высокоинтеллектуальных" средств очертил их область применения, их преимущества и недостатки. В частности, выделилась ниша широкого класса задач, не требующих глубокого логического вывода, не предъявляющих высоких требований к пользователю, но имеющих широкое распространение во многих сферах деятельности. Эти задачи можно назвать поиском информации в разнородном информационном пространстве. Человека окружает множество различных баз данных, доступных ему, и, в принципе, относящимся к решаемой задаче. Но каждая из баз имеет свой язык, свои средства доступа; базы данных делались не для текущей задачи пользователя, а для других целей. Как их эффективно использовать? Для удовлетворения этой потребности возник даже специальный слой посредников - информационные менеджеры, и многие Западные университеты готовят таких специалистов. Именно эти специалисты трансформируют информационную потребность пользователя в запросы к конкретным базам данных и обобщают полученную информацию. Можно ли сделать эту работу в автоматизированном варианте? Ответом на этот вопрос и разработкой соответствующих технологий и заняты специалисты по "интеллектуальным" технологиям поиска информации, который во многом базируется на теории нечетких систем.

Во - вторых, развитие информационного пространства претерпело в последнее время кардинальные изменения. В практическую деятельность давно вошли такие действительно всемирные и глобальные сети как INTERNET, которые переворачивают обычную технологию информационной работы. Без интеллектуальных средств поиска информации, ее доставки и анализа работа становится практически не возможной. Учитывая большую неопределенность запросов, неопределенность в местонахождении информации и ее объемы, данные средства также во многом являются "нечеткими". Важным также является широкое развитие корпоративных "хранилищ данных" [24] и технологий их эффктивного использования для анализа и повышения эффективности функционирования различных бизнес-процессов крупных и средних компаний [6, 16]. Такие средства, как business intelligents [1, 27], во многом оперируют с нечеткой, обобщенной информацией.

И, наконец, актуальным становится вопрос эффективного

Таблица 1.1 Возможные варианты нечеткости в среде поиска информации

| Запрос | База данных |
|---|---|
| четкий | четкая |
| четкий | нечеткая |
| нечеткий | четкая |
| нечеткий | нечеткая |

использования накопленных информационных ресурсов. Многие базы данных разрабатывались 20 - 30 и более лет назад, накопленные ресурсы представляют значительную ценность, но из-за устаревших средств поиска и обработки информации их использование затруднено. Перевод существующих баз данных на новую технико - программную базу часто является очень дорогостоящим проектом, поэтому создание средств поиска информации, позволяющих работать с такими базами данных на уровне современных технологий и не затрагивающих собственно сами базы является экономически оправданной стратегией повышения эффективности использования имеющихся информационных ресурсов.

## 1.2. Классификация задач поиска информации в нечеткой среде

Под средой поиска информации будем понимать пару ⟨запрос, база данных ⟩. С точки зрения четкости - нечеткости могут быть четыре варианта (Таблица 1.2).

Под *четким запросом* понимается логическое высказывание, термы которого выражаются обычными средствами теории множеств. Это означает, что можем либо перечислить значения признаков интересующих нас объектов, либо указать границы изменения параметров признаков и связать данные пары ”признак - значение” логическими связками. Для многих задач этого оказывается достаточно для удовлетворения информационной потребности.

Под *четкими базами данных* в нашем случае понимается совокупность записей, значения атрибутов которых есть либо строковые значения, однозначно понимаемые пользователями (названия объектов, марки и т.п.), либо численные значения.

*Нечеткий запрос* в отличии от четкого может содержать термы с нечеткими значениями. Например, значением признака "Размер" могут быть "Большой", "Не большой и не маленький"; значением признака "Возраст" могут быть "Молодой", "Более или менее молодой"; значением признака "Надежность" - "Высокая", "Удовлетворительная" и т.п.

Такого же типа разница заключается между четкими и *нечеткими базами данных*: атрибуты последних могут иметь нечеткие значения.

Как видно из таблицы 1.2, наиболее общей является ситуация 4. Ситуации 2 и 3 являются ее частными случаями, и, кроме того, являются симметричными. Ситуация 1 является наиболее простой и изученной. Практически все базы данных и все связанные с поиском информации научные результаты принадлежат именно этой ситуации, поэтому мы не будем ее описывать. Учитывая вышесказанное, приведем сначала результаты исследований ситуации 4 (раздел 3), а затем ситуации 3 (раздел 4).

Для этого нам понадобится формализовано описывать множества значений качественных признаков (атрибутов). Моделью таких структур является семантическое пространство, которое можно рассматривать как частный случай понятия лингвистической переменной, введенной Заде в [10].

## Глава 2.

# Семантические пространства и их свойства

Прежде всего, напомним понятие лингвистической переменной.

## 2.1. Понятие лингвистической переменной

Опираясь на понятие нечеткого множества Заде в [10] вводит понятие нечеткой переменной как тройки

$$\langle \alpha, U, G \rangle,$$

где $\alpha$ - наименование (имя) нечеткой переменной;

$U$ - область ее определения (универсальное множество);

$G$ - нечеткое множество в $U$, описывающее ограничения на возможные значения нечеткой переменной $\alpha$ (ее семантику).

В зависимости от характера множества $U$ нечеткие переменные могут быть разделены на числовые и нечисловые. К числовым относятся нечеткие переменные, у которых $U \subset R^1$.

Дальнейшим шагом является введение понятия лингвистической переменной как пятерки

$$\langle A, T(A), U, V, M \rangle,$$

где $A$ - название переменной;

$T(A)$ - терм-множества переменной $A$, т.е. множество названий лингвистических значений переменной $A$, причем каждое из таких значений - нечеткая переменная со значениями из универсального множества $U$;

$V$ - синтаксическое правило (обычно грамматика), порождающее названия значений лингвистической переменной $A$;

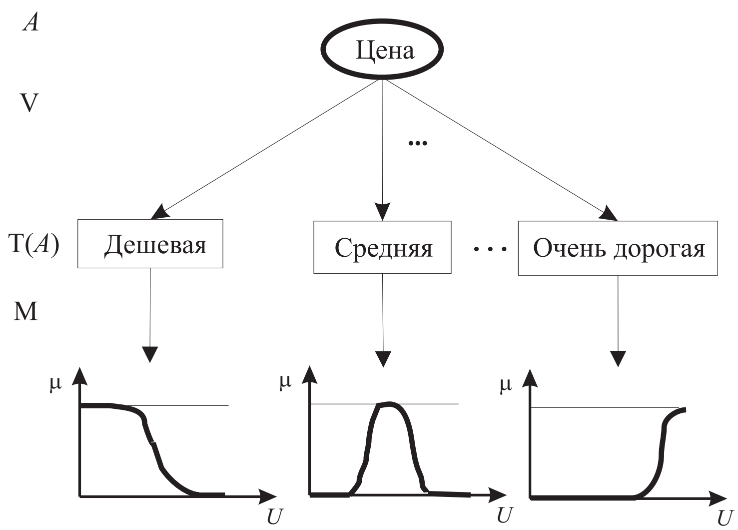


Рис. 2.1 Графическое изображение лингвистической переменной "Цена".

$M$ - семантическое правило, которое ставит в соответствие каждой нечеткой переменной из $T(A)$ нечеткое подмножество универсального множества $U$.

Пример лингвистической переменной "Цена" представлен на рис. 2.1

Заде различает *базовые термины* (молодой, среднего возраста, пожилой, ...) и *модификаторы* (очень, не-, слегка, ...). Модификаторы могут применяться как к базовым терминам (очень молодой, не старый, ...), так и к комбинациям базового термина и модификатора (очень-очень старый, слегка не молодой, ...). Правила применения модификаторов задаются синтактическим правилом $V$.

Разница между базовыми терминами и модификаторами заключается в следующем. Для базовых терминов функции принадлежности задаются, а модификаторы действуют как некоторые операторы над этими функциями. Например, в качестве "очень" предлагается следующая модификация функции принадлежности термина "молодой" [10]: $\mu_{\text{очень молодой}}(u) =$

$\mu^2_{\text{молодой}}(u)$. Аналогично $\mu_{\text{слегка молодой}}(u) = \mu^{1/2}_{\text{молодой}}(u)$ и т.п. Вопросы адекватности таких преобразовний практически не изучались. Одна из причин этого заключается в большой неопределенности операций: для разных ситуаций могут быть разные результаты. Проблема заключается в том, что для разных контекстов функции принадлежности одного и того же термина могут быть разными. Эта проблема изучается в рамках концепции семантического пространства. Более подробно с теорией и приложениями лингвистической переиенной можно ознакомиться в [10], мы же более подробно остановимся на концепции семантического пространства, его свойствах и приложениях.

## 2.2. Концепция полного ортогонального семантического пространства (ПОСП)

Рассмотим ситуацию, когда лингвистическая переменная $A =$ "РОСТ" имеет два терм-множества $T_1(A) = \{$низкий, высокий$\}$ и $T_2(A) = \{$низкий, средний , высокий$\}$.

Интуитивно ясно, что функции принадлежности понятий "низкий" и "высокий" в первом и во втором случае будут различаться: новое понятие "средний" модифицирует их, сдвигает к концам универсума (Рис. 2.2)

Последнее говорит о том, что семантика некоторого термина зависит от контекста, или набора значений соответствующей лингвистической переменной. Таким образом, функция принадлежности любого термина без указания контекста, вообще говоря, не имеет смысла.

Будем называть семантическим пространством лингвистическую переменную с фиксированным терм-множеством, т.е. четверку

$$S = \langle A, T(A), U, M \rangle.$$

Иными словами, семантическое пространство - это набор нечетких переменных

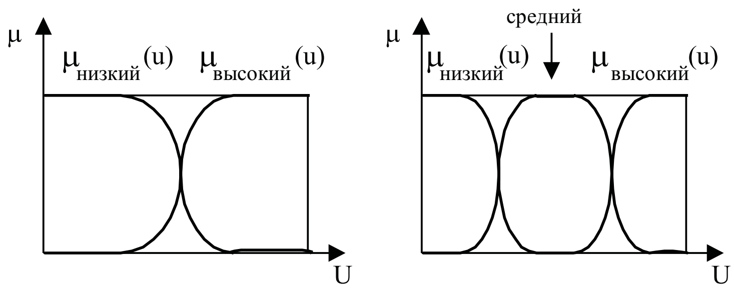


Рис. 2.2 Модификация функций принадлежности понятий.

$$S = \langle \alpha_1, U, G_1 \rangle, \ldots, \langle \alpha_n, U, G_n \rangle. \qquad (2.1)$$

При этом для одного и того же имени $A$ могут существовать различные пространства

$$S_1 = \langle A, T_1(A), U, M_1 \rangle, \ldots, S_k = \langle A, T_k(A), U, M_k \rangle.$$

Можно ли как-то сравнивать семантические пространства, выбирать наилучшее в некоторм смысле? Рассмотрим следующий пример.

**Пример 1** *Рассмотрим процесс описания человеком некоторых реальных объектов на примере описания других людей. Описывая ВОЗРАСТ человека, мы можем использовать несколько вариантов множества значений признака "ВОЗРАСТ":*

- $T_1$ = *{молодой, старый };*
- $T_2$ = *{молодой, среднего возраста, старый };*

$\vdots$

- $T_n$ = *{юный, очень молодой, ..., очень старый }.*

*Какое из этих множеств лучше с точки зрения "легкости" описания возраста?*

*Множество* $T_1$ *не является таковым, так как существует много людей, для которых обазначения одинаково не подходят.*

*Мы испытывам трудности описания из-за недостатка значений.*

*Множество $T_n$ также является "плохим" из-за того, что для одного и того же реального объекта могут оказаться одинаково подходящими несколько значений признака.*

Вопросам оценки степени неопределенности процесса описания реальных объектов посвящен следующий раздел.

Для возможности оценки степени неопределенности необходимо сформулировать некоторые требования для функций принадлежности используемых понятий и их совокупностей, образующих семантическое пространство. Таким образом, ниже мы будем рассматривать не все возможные семантические пространства, а некоторое их подмножество.

При формулировке таких требований необходимо удовлетворить двум противоречивым критериям:

- требования должны быть достаточно "мягкими", чтобы получившееся подмножество было достаточно широким и включало в себя большинство практических ситуаций;

- требования должны быть достаточно "жесткими", чтобы давать возможность формального введения различных понятий и изучения их свойств.

Итак, рассмотрим семантическое пространство ( 2.1). В качестве его модели рассмотрим совокупность $t$ функций принадлежности, заданных на одном универсальнм множестве $U$. Для краткости будем обозначать такую совокупность $s_t$.

Будем считать, что функции принадлежности $s_t$ определены на некотором отрезке $U \subseteq R^1$ и удовлетворяют следующим требованиям:

(1) нормальность [14]: $\forall j(1 \leq j \leq t) \quad \exists U_j^1 \neq \emptyset$, где $U_j^1 = \{u \in U : \quad \mu_j(u) = 1\}$, $U_j^1$ является отрезком;

(2) $\mu_j(u)$ не убывает слева от $U_j^1$ и не возрастает справа от $U_j^1$.

Данные ограничения являются довольно естественными для функций принадлежности понятий, образующих семантическое пространство. Действительно, (1) означает, что для каждого понятия существует хотя бы один объект, являющийся для него типичным; (2) может быть интерпретировано как требование плавности, мягкости границ используемых понятий.

В будущем нам понадобится использование наряду с функциями принадлежности и характеристических функций, поэтому к сформулированным требованиям добавим требование

(3) функции не могут иметь более двух точек разрыва первого рода.

Обозначим через $L$ множество функций, удовлетворяющих требованиям (1) - (3).

Сформулируем также требования на совокупности функций из $L$, образующих множество $s_t$. Будем считать, что множество из $t$ таких функций удовлетворяет следующим двум требованиям:

(4) полнота: $\forall u \in U \quad \exists j (1 \leq j \leq t): \quad \mu_j(u) \neq 0$;

(5) ортогональность: $\forall u \in U \quad \sum\limits_{j=1}^{t} \mu_j(u) = 1$.

Эти ограничения также являются довольно естественными. Требование (4) означает, что для каждого объекта найдется хотя бы одно понятие, его описывающее с ненулевой степенью; (5) означает достаточную разделимость понятий, образующих семантическое пространство, отсутствие синонимии или семантически близких терминов.

Будем обозначать через $G_t(L)$ множество из $t$ функций из $L$, удовлетворяющих требованиям (4), (5).

Семантичесие пространства, функции принадлежности понятий которых принадлежат $G_t(L)$, будем называть полными ортогональными семантическими пространствами (ПОСП).

В рамках ПОСП можно ввести понятие степени нечеткости или меры внутренней неопределенности, которое позволяет выбирать наилучшие пространства для описания человеком реальных объектов (см. пример 1).

## 2.3. Степень нечеткости ПОСП

Прежде всего заметим, что определенное в 2.2 множество $L$ функций является подмножеством множества интегрируемых на отрезке функций, поэтому мы можем ввести метрику на $L$, например,

$$\rho(f,g)=\int_U |f(u)-g(u)|du, \quad f\in L,\ g\in L.$$

Мы также можем ввести метрику в $G_t(L)$.

**Лемма 1** *Пусть* $s_t\in G_t(L),\ s'_t\in G_t(L)$,
$s_t=\{\mu_1(u),\mu_2 u(),\dots,\mu_t(u)\}, \quad s'_t=\{\mu'_1(u),\mu'_2(u),\dots,\mu'_t(u)\}$,
$\rho(f,g)$ *- некоторая метрика в* $L$.

*Тогда*

$$d(s_t,s'_t)=\sum_{j=1}^{t}\rho(\mu_j,\mu'_j) \tag{2.2}$$

*есть метрика в* $G_t(L)$.

**Доказательство** леммы приведено в приложении А.1.

Для формулировки аксиом необходимо определить совокупность множеств, базирующуюся на данной совокупности нечетких множеств и являющихся "четкими". Это множество характеристических функций, определяемых следующим образом.

Пусть $s_t\in G_t(L)$ определена на $U$ и включает функции принадлежности $\mu_1(u),\mu_2(u),\dots,\mu_t(u)$. Построим совокупность характеристических функций $\tilde{s_t}$, состоящую из функций $h_1(u),h_2(u),\dots$,
$h_t(u)$, где

$$h_i(u)=\begin{cases} 1, & \text{если} \quad \max_{1\le j\le t}\mu_j(u)=\mu_i(u)\\ 0, & \text{в противном случае}\end{cases} \tag{2.3}$$

Будем называть $\tilde{s_t}$ ближайшей совокупностью характеристических функций для $s_t\in G_t(L)$.

Под степенью нечеткости $s_t\in G_t(L)$ будем понимать значение функционала $\xi(s_t)$, определенного на множестве функций принадлежности $s_t$ и удовлетворяющего следующим аксиомам:

$$\begin{array}{lll}
A1. & 0\le\xi(s_t)\le 1 & \forall s_t\in G_t(L).\\
A2. & \xi(s_t)=0\Leftrightarrow & \forall u\in U\ \exists j(1\le j\le t):\\
 & & \mu_j(u)=1,\ \mu_i(u)=0\ \forall i\ne j.\\
A3. & \xi(s_t)=1\Leftrightarrow & \forall u\in U\ \exists i_1,i_2(1\le i_1,i_2\le t):\\
 & & \mu_{i_1}(u)=\mu_{i_2}(u)=\max_{1\le j\le t}\mu_j(u).
\end{array}$$

$A4$. Пусть $s_t$ и $s'_{t'}$ определены на универсальных множествах $U$ и $U'$ соответственно; $t$ и $t'$ могут быть равны или не равны друг другу. Тогда

$\xi(s_t) \leq \xi(s'_{t'})$, если $\rho(s_t, \tilde{s_t}) \leq \rho(s'_{t'}, \tilde{s_{t'}}')$, где $\rho(.,.)$ - некоторая метрика в $G_t(L)$.

Аксиома $A1$ определяет границы изменения степени нечеткости.

Аксомы $A2$ и $A3$ описывают совокупности нечетких множеств, на которых $\xi(s_t)$ достигает максимальные и минимальные значения, то есть максимально "четкие" и максимально "нечеткие" совокупности нечетких множеств соответственно.

Аксиома $A4$ определяет для каждой пары совокупнстей нечетких множеств правило сравнения степени их нечеткости. Ее можно интерпретировать следующим образом: чем ближе некоторая совокупность нечетких множеств к своей ближайшей совокупности характеристических функций, тем меньше степень ее нечеткости.

Существуют ли функционалы, удовлетворяющие сформулированным аксиомам? Ответ на этот вопрос дает следующая

**Теорема 1** (Теорема существования). *Пусть $s_t \in G_t(L)$. Тогда функционал*

$$\xi(s_t) = \frac{1}{|U|} \int_U f(\mu_{i_1^*}(u) - \mu_{i_2^*}(u))du, \tag{2.4}$$

*где*

$$\mu_{i_1^*}(u) = \max_{1 \leq j \leq t} \mu_j(u), \quad \mu_{i_2^*}(u) = \max_{1 \leq j \leq t;\ j \neq i_1^*} \mu_j(u), \tag{2.5}$$

*$f$ удовлетворяет следующим требованиям:*

$F1: \ f(0) = 1, \quad f(1) = 0$;

$F2: \ f$ *убывает,*

*- является степенью нечеткости $s_t$, то есть удовлетворяет аксиомам $A1 - A4$.*

**Доказательство** теоремы приведено в приложении A.2.

Достаточно очевидно, что существует только одна линейная функция, удовлетворяющая $F1$, $F2$. Это функция

$$f(x) = 1 - x.$$

Можно также описать подмножество полиномов второй степени, удовлетворяющих $F1,\ F2$. Это параметрическое семейство функций

$$f_a(x) = ax^2 - (1+a)x + 1.$$

Подмножества функций других типов (логарифмических, тригонометрических и др.), удовлетворяющих $F1,\ F2$ моут быть описаны аналогичным образом. Подставляя эти функции в формулу ( 2.4), мы получаем функционалы, удовлетворяющие $A1 - A4$, то есть степени нечеткости.

Какие из этих классов функционалов "лучше"? Это довольно сложный вопрос, ответ на который зависит от конкретного приложения. Мы не будем углубляться в конкретные проблемы, а изучим некоторые общие свойства простейшего из таких функционалов - функционала из класса линейных функций $f$.

## 2.4. Некоторые свойства степени нечеткости ПОСП

Мы приведем свойства степени нечеткости для линейной функции $f$:

$$\xi(s_t) = \frac{1}{|U|}\int_U (1 - (\mu_{i_1^*}(u) - \mu_{i_2^*}(u)))du, \tag{2.6}$$

где $\mu_{i_1^*}(u),\ \mu_{i_2^*}(u)$ описываются ( 2.5).

Функционал $\xi(s_t)$ может быть интерпретирован как средняя степень трудностей описания человеком реальных объектов (ситуаций) в рамках соответствующего семантического пространства.

*Интерпретация.* Рассмотрим процесс описания человеком реальных объектов. Мы не имеем никакой неопределенности при лингвистическом описании объекта, имеющего "физическое" значение признака $u_1$ (Рис. 2.3). Мы присвоим ему лингвистическое значение $a_1$ без сомнений и колебаний.

Мы можем повторить данные рассуждения для объекта, имеющего "физическое" значение признака $u_5$. Мы без колебаний выбираем термин $a_2$ для его лингвистического описания

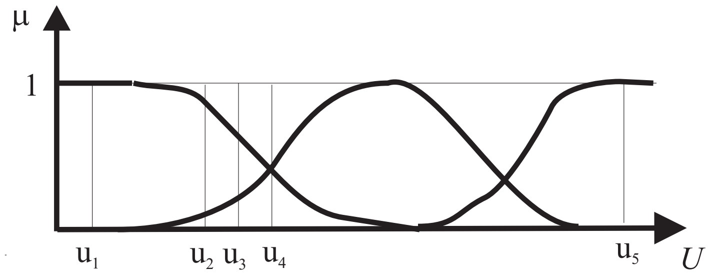


Рис. 2.3 Модель лингвистического описания объектов.

без сомнений. Мы начинаем испытывать трудности при выборе лингвистического значения для объекта, имеющего "физическое" значение признака $u_2$. Эти трудности возрастают ($u_3$) и достигают максимального значения при описании объектов, имеющих "физическое" значение признака $u_4$: для таких объектов оба лингвистических значения одинаково подходят.

Если мы рассмотрим значения подинтегральной функции

$$\eta(s_t, u) = 1 - (\mu_{i_1^*}(u) - \mu_{i_2^*}(u))$$

в этих точках, мы можем увидеть, что

$$0 = \eta(s_t, u_5) = \eta(s_t, u_1) < \eta(s_t, u_2) < \eta(s_t, u_3) < \eta(s_t, u_4) = 1.$$

Таким образом, значение интеграла ( 2.6) мы действительно можем интерпретировать как среднию степень трудностей описания человеком реальных объектов (ситуаций) в рамках соответствующего ПОСП.

Рассмотрим некоторые свойства функционала ( 2.6). Для этого рассмотрим следующие подмножества $L$:

$\overline{L}$ - множество кусочно - линейных функций из $L$, которые являются линейными на множестве неопределенности

$$\overline{U} = \{u \in U : \forall j \ (1 \leq j \leq t) \ 0 < \mu_j(u) < 1\},$$

$\widehat{L}$ - множество функций из $L$, являющихся кусочно-линейными на $U$ (включая $\overline{U}$).

**Теорема 2** *Пусть* $s_t \in G_t(\overline{L})$*. Тогда* $\xi(s_t) = \frac{d}{2|U|}$*, где* $d = |\overline{U}|$*.*

**Доказательство** теоремы приведено в приложении A.3.

**Теорема 3** *Пусть* $s_t \in G_t(\widehat{L})$*. Тогда*

$$\xi(s_t) = c\frac{d}{|U|}, \tag{2.7}$$

*где* $d = |\overline{U}|, c < 1,\ c = Const.$

**Доказательство** теоремы достаточно очевидно.

Так как любая $s_t \in G_t(L)$ может быть со сколь угодно большой точностью аппроксимирована совокупностью нечетких множеств из $s_t \in G_t(\widehat{L})$, то соотношение ( 2.7) справедливо для всех $s_t \in G_t(L)$.

Пусть $g$ некоторая взаимно-однозначная функция, определенная на $U$. Эта функция индукцирует преобразование некоторой $s_t \in G_t(L)$, определенной на универсальном множестве $U$ в $g(s_t)$, определенной на универсальном множестве $U'$, где

$U' = g(U) = \{u' :\ u' = g(u),\ u \in U\}$.

Это преобразование можно определить следующим образом: $g(s_t)$ есть множество функций принадлежности

$$\{\mu'_1(u'), ..., \mu'_t(u')\},$$

где $\mu'_j(u') = \mu'_j(g(u)) = \mu_j(g^{-1}(u')) = \mu_j(u), \quad \mu_j(u) \in s_t,$
$1 \le j \le t.$

Следующий пример иллюстрирует данное определение.

**Пример 2** *Пусть* $s_t \in G_t(L)$*,* $U$ *- универсум* $s_t$ *и* $g$ *- растяжение (сжатие) универсума* $U$*. В этом случае* $g(s_t)$ *есть совокупность функций принадлежности, полученная из* $s_t$ *таким же растяжением (сжатием).*

**Теорема 4** *Пусть* $s_t \in G_t(L)$*,* $U$ *- универсум* $s_t$*,* $g$ *- некоторая линейная взаимно-однозначная функция на* $U$ *и* $\xi(s_t) \neq 0$*. Тогда* $\xi(s_t) = \xi(g(s_t))$*.*

**Доказательство** теоремы приведено в приложении A.4.

Это свойство означает, что человек описывает разнотипные объекты в рамках некоторогосемантического пространства с равными трудностями, если физические параметры объектов одного типа можно получить из параметров объектов другого типа некоторым линейным преобразованием. Например, используя множество термов {высокий, средний, низкий} мы описываем людей, деревья, здания с одинаковыми трудностями; используя множество значений {очень близко, близко, не близко, далеко} мы описываем расстояния между модекулами, улицами в городе, городами на карте и т.п. с одинаковыми трудностями.

Степень нечеткости одного множества, индуцированная $\xi(s_t)$ модет быть определена как степень нечеткости тривиальной совокупности нечетких множеств, определенной одним множеством $\mu(u)$:

$$\xi(\mu) = \frac{1}{|U|} \int_U (1 - |2\mu(u) - 1|) du. \tag{2.8}$$

Не трудно доказать, что ( 2.8) обладает всеми свойствами степени нечеткости множества, изложенными в [18].

**Глава 3.**

# Поиск по нечетким запросам в нечетких базах данных

В данном разделе рассматриваются некоторые вопросы приложений теории нечетких множеств в информационных системах. В частности, рассматривается процесс поиска информации в базах данных, для которых источником информации является человек. Определяются параметры качества поиска информации в таких базах данных (средние индивидуальные потери информации и шумы) и устанавливается их связь со степенью нечеткости соответствующих семантических пространств. На основе такой связи формулируется методика выбора оптимального с точки зрения качества поиска информации описания объектов. Рассматриваются также вопросы построения интерфейса к (обычным) базам данных, позволяющего вводить и определять понятия пользователя, а также осуществлять по ним поиск информации.

## 3.1. Лингвистические базы данных

Рассмотрим следующую модель (Рис. 3.1).

Наблюдая объекты предметной области системы, человек описывает их для базы данных. Пусть человек - источник информации не может пользоваться измерительными приборами, то есть описание объекта есть набор лингвистических значений определенных характеристик или признаков. Такие описания образуют базу данных информационной системы.

Пользователями системы является также некоторое множество людей. Решая свои информационные задачи, пользователи используют базу данных как информационную модель предметной области. При этом качество решения таких задач во многом зависит от адекватности информационной модели (то есть базы данных) предметной области в представлении кон-


Рис. 3.1 Модель лингвистической базы данных.

кретного пользователя. Содержательной интерпретацией такой адекватности является степень согласования мнений источника информации и пользователя относительно значений характеристик лингвистических описаний объектов.

Первый вопрос, который возникает при анализе модели при приведенной выше содержательной интерпретации адекватности, может быть сформулирован следующим образом. Можно ли, учитывая некоторые законы и модели восприятия человеком объектов реального мира, сформулировать такое правило выбора множества значений качественного признака, что человек будет испытывать минимально возможную неопределенность при описании конкретного объекта в рамках данного множества значений признака?

Другой формулировкой этого же вопроса может быть следующая. Можно ли предложить такое правило выбора множества значений качественного признака, в рамках которого разные люди будут описывать объекты с наименьшей степенью рассогласованности? Заметим, что ответ на этот вопрос есть и он получен нами при изучении свойств степени нечеткости полных ортогональных семантических пространств ( 2.3). Не менее важным является вопрос практического использования полученного результата, а именно: если мы будем описывать объекты с минимальной степенью неопределенности, что это даст в плане повышения качества поиска информации в базах данных? Если использование оптимального множества значений не оказывает существенного влияния на повышение качества поиска информации, то его использование не имеет практического смысла. Ответ на этот вопрос дается в 3.2. Там же приводится методика выбора оптимального множества значений качественного признака с точки зрения качества поиска информации в лингвистических базах данных. Важным для практического использования также является изучение проблемы устойчивости полученной методики. Если малые изменения исходных данных могут оказать существенное влияние на результаты, получаемые на основе методики, то ее практическое использование является весьма сомнительным. Эта проблема рассматривается в 3.3.1.

## 3.2. Потери информации и шумы

При задач поиска информации пользователь задает системе некоторый запрос, например: "Выдать описания всех объектов, значение признака "Размеры" которых равно "Маленький"". Могут быть более сложные запросы, состоящие из комбинаций элементарных запросов, аналогичных приведенному, и связок "И", "ИЛИ", "НЕ". В ответ на запрос из базы данных выдается некоторое количество описаний объектов, ему удовлетворяющих. Представим теперь, что пользователь имеет возможность анализировать не только лингвистические описания объектов, но и сами объекты. При этом могут возникнуть следующие ситуации:

- часть выданных на запрос объектов пользователь может посчитать не соответствующими запросу. Это те объекты, которые описаны источником информации как "маленькие", а с точки зрения пользователя они таковыми не являются (а являются, например, "средними"). Такие описания будем называть информационным шумом;

- если бы пользователь мог просмотреть все объекты, описания которых хранятся в системе, он, возможно, нашел бы такие объекты, которые, по его мнению, соответсвуют запросу, но не были выданы системой. Это те объекты, которые описаны источником как, например, "средний", но с точки зрения пользователя они являются "маленькими". Такие описания будем называть потери информации.

Подсчитаем объемы потерь информации и информационных шумов.

Пусть $I(O_1), ..., I(O_q)$ - описания объектов $O_1, ..., O_q$, хранящихся в системе. $I(O_j)(j = 1, ..., q)$ - набор значений качественных признаков. Рассмотрим некоторый признак. Он имеет множество значений $X = \{a_1, ..., a_t\}$. Математической моделью множества значений качественного признака является совокупность функций принадлежности $s_t \in G_2(L)$. Будем обозначать элементарный запрос "Выдать описания объектов, имеющих значение признака, равное $a_j$" через $\langle I(O) = a_j \rangle$.

Рассмотрим частный случай $t = 2$.

Рассмотрим некоторую точку $u^\star \in U$. Введем следующие обозначения:

$N(u^\star)$ - количество описаний объектов в базе данных системы, имеющих физическое значение признака, равное $u^\star$;

$N^E$ - количество пользователей информационной системы.

Тогда $N_{a_1}(u^\star) = \mu_{a_1}(u^\star)N(u^\star)$ - количество описаний объектов в базе данных, имеющих физическое значений признака $u^\star$, и описанных источником информации как $a_1$;

$N_{a_2}(u^\star) = \mu_{a_2}(u^\star)N(u^\star)$ - количество описаний объектов в базе данных, имеющих физическое значений признака $u^\star$, и описанных источником информации как $a_2$;

$N^E_{a_1}(u^\star) = \mu_{a_1}(u^\star)N^E$ - количество пользователей системы, считающих, что $u^\star$ есть $a_1$;

$N^E_{a_2}(u^\star) = \mu_{a_2}(u^\star)N^E$ - количество пользователей системы, считающих $u^\star$ $a_2$.

Заметим, что, так как $s_t \in G_2(L)$, то

$$N_{a_1}(u^\star) + N_{a_2}(u^\star) = \mu_{a_1}(u^\star)N(u^\star) + \mu_{a_2}(u^\star)N(u^\star) = N(u^\star),$$

$$N^E_{a_1}(u^\star) + N^E_{a_2}(u^\star) = \mu_{a_1}(u^\star)N^E + \mu_{a_2}(u^\star)N^E = N^E.$$

При запросе $\langle I(O) = a_1 \rangle$ выдается $N_{a_1}(u^\star)$ описаний, имеющих физическое значение признака $u^\star$. При этом $N^E_{a_1}(u^\star)$ пользователей несут потери в объеме $N_{a_2}(u^\star)$ описаний. Это описания тех объектов в базе данных, которые имеют физическое значение признака $u^\star$ и описаны источником информации как $a_2$. Оставшиеся $N^E_{a_2}(u^\star)$ пользователей имеют информационный шум в объеме выдавшихся $N_{a_1}(u^\star)$ описаний объектов.

Таким образом, средние индивидуальные потери информации, возникающие при поиске по данному запросу для объектов, имеющих физическое значение признака, равное $u^\star$

$$\varphi_{a_1}(u^\star) = \frac{1}{N^E} N^E_{a_1}(u^\star)N(u^\star) = \frac{1}{N^E}\mu_{a_1}(u^\star)N^E\mu_{a_2}(u^\star)N(u^\star). \tag{3.1}$$

Средние индивидуальные шумы в этом случае

$$\psi_{a_1}(u^\star) = \frac{1}{N^E} N^E_{a_2}(u^\star)N(u^\star) = \frac{1}{N^E}\mu_{a_2}(u^\star)N^E\mu_{a_1}(u^\star)N(u^\star). \tag{3.2}$$

Как видно из формул ( 3.1) и ( 3.2),

$$\varphi_{a_1}(u^\star) = \psi_{a_1}(u^\star).$$

Это является следствием того, что $s_t \in G_2(L)$, а также того, что мы в рамках исследуемой модели не различаем функции принадлежности источника информации и пользователя. Последнее будет изучено нами в рамках расширения модели (раздел  3.3.1).

Средние индивидуальные потери информации и шумы, возникающие при поиске по данному запросу для всех объектов $\Phi_{a_1}(U)$ и $\Psi_{a_1}(U)$ соответственно, определяются как

$$\Phi_{a_1}(U) = \frac{1}{|U|}\int_U \varphi_{a_1}(u)du, \tag{3.3}$$

$$\Psi_{a_1}(U) = \frac{1}{|U|}\int_U \psi_{a_1}(u)du. \tag{3.4}$$

Не трудно видеть, что

$$\Phi_{a_1}(U) = \Psi_{a_1}(U) = \frac{1}{|U|}\int_U \mu_{a_1}(u)\mu_{a_2}(u)N(u)du \tag{3.5}$$

Повторяя приведенные рассуждения для запроса $\langle I(O) = a_2\rangle$, получим, что и в этом случае средние индивидуальные потери информации и шумы равны друг другу ($\Phi_{a_2}(U) = \Psi_{a_2}(U)$) и равны правой части ( 3.5).

Рассмотрим теперь все множество значений поискового признака $X = \{a_1, a_2\}$. Средние индивидуальные потери информации и шумы при поиске информации по нему ($\Phi_X(U)$ и $\Psi_X(U)$ соответственно) естественно определить как

$$\Phi_X(U) = p_1\Phi_{a_1}(U) + p_2\Phi_{a_2}(U), \tag{3.6}$$

$$\Psi_X(U) = p_1\Psi_{a_1}(U) + p_2\Psi_{a_2}(U), \tag{3.7}$$

где $p_i (i = 1, 2)$ - вероятность запроса по $i$ - значению признака.

Подставляя ( 3.5) в ( 3.6), ( 3.7) с учетом сделанного выше замечания для запроса $\langle I(O) = a_2\rangle$ и очевидного соотношения $p_1 + p_2 = 1$, получаем:

$$\Phi_X(U) = \Psi_X(U) = \frac{1}{|U|}\int_U \mu_{a_1}(u)\mu_{a_2}(u)N(u)du \tag{3.8}$$

Полученные результаты довольно естественно обобщаются на случай $t > 2$.

Действительно, в этом случае область интегрироания можно представитьв виде

$$U = U_1 \cup U_{12} \cup U_2 \cup ... \cup U_{t-1,t} \cup U_t, \tag{3.9}$$

где $U_j = [\overline{u}_{j-1,R}, \overline{u}_{j+1,L}](j = 2, ..., t; \overline{u}_{0,R} = \overline{u}_{1,L}, \overline{u}_{t+1,L} = \overline{u}_{t,R})$ - подмножество $U$, на котором $\mu_{a_j}(u) = 1$, и, в силу свойства ортогональности $G_2(L)$ $\mu_{a_i}(u) = 0 \forall i \neq j$.

$U_{j-1,j} = [\overline{u}_{j,L}, \overline{u}_{j-1,R}](j = 2, ..., t)$ - подмножество $U$, на котором $0 < \mu_{a_{j-1}}(u), \mu_{a_j}(u) < 1; \mu_{a_i}(u) = 0$ для $i \neq j-1, i \neq j$.

Рассмтрим запрос $\langle I(O) = a_j \rangle (1 \leq j \leq t)$. В этом случае на качество поиска нужных объектов оказывают влияние соседние значения признака: левое $(j-1)$ и правое $(j+1)$. Для средних потерь информации и шумов, таким образом, справедливо:

$$\Phi_{a_j}(U) = \Phi_{a_j}^{a_{j-1}}(U) + \Phi_{a_j}^{a_{j+1}}(U), \tag{3.10}$$

$$\Psi_{a_j}(U) = \Psi_{a_j}^{a_{j-1}}(U) + \Psi_{a_j}^{a_{j+1}}(U), \tag{3.11}$$

где

$$\begin{aligned} \Phi_{a_j}^{a_{j-1}}(U) = \Psi_{a_j}^{a_{j-1}}(U) &= \frac{1}{|U|} \int_U \mu_{a_j}(u) \mu_{a_{j-1}}(u) N(u) du = \\ &= \frac{1}{|U|} \int_{U_{j-1,j}} \mu_{a_{j-1}}(u) \mu_{a_j}(u) N(u) du, \end{aligned} \tag{3.12}$$

$$\begin{aligned} \Phi_{a_j}^{a_{j+1}}(U) = \Psi_{a_j}^{a_{j+1}}(U) &= \frac{1}{|U|} \int_U \mu_{a_j}(u) \mu_{a_{j+1}}(u) N(u) du = \\ &= \frac{1}{|U|} \int_{U_{j,j+1}} \mu_{a_j}(u) \mu_{a_{j+1}}(u) N(u) du \end{aligned} \tag{3.13}$$

Первое и второе равенства в ( 3.12), ( 3.13) можно получить, повторив рассуждения вывода формулы ( 3.8), считая для ( 3.12) $j = 2$ и для ( 3.13) $j = 1$. Последние соотношения в ( 3.12), ( 3.13) следуют из определений множеств $U_{j-1,j}$ и $U_{j,j+1}$.

В этом случае математическое ожидание средних индивидуальных потерь информации и информационных шумов, возникающих при поиске информации по данному признаку, будут равны соответственно

$$\Phi_X(U) = \sum_{j=1}^{t} p_j \Phi_{a_j}(U), \tag{3.14}$$

$$\Psi_X(U) = \sum_{j=1}^{t} p_j \Psi_{a_j}(U), \tag{3.15}$$

где $p_j$ - вероятность (частота) запроса по $j$ - значению признака $X$, $\sum_{j=1}^{t} p_j = 1$.

Заметим, что для крайних значений $j = 1$ и $j = t$

$$\begin{aligned} \Phi_{a_1}(U) = \Phi_{a_1}^{a_2}(U) = \Psi_{a_1}(U) = \Psi_{a_1}^{a_2}(U) = \\ = \frac{1}{|U|} \int_{U_{12}} \mu_{a_1}(u)\mu_{a_2}(u)N(u)du, \end{aligned}$$

$$\begin{aligned} \Phi_{a_t}(U) &= \Phi_{a_t}^{a_{t-1}}(U) = \Psi_{a_t}(U) = \Psi_{a_t}^{a_{t-1}}(U) = \\ &= \frac{1}{|U|} \int_{U_{t-1,t}} \mu_{a_{t-1}}(u)\mu_{a_t}(u)N(u)du. \end{aligned}$$

Подставляя ( 3.10) - ( 3.13) в ( 3.14), ( 3.15) с учетом сделанного замечания, получим:

$$\begin{aligned} \Phi_X(U) &= \Psi_X(U) = p_1 \frac{1}{|U|} \int_{U_{12}} \mu_{a_1}(u)\mu_{a_2}(u)N(u)du + \\ &+ \sum_{j=2}^{t-1} p_j \frac{1}{|U|} \Big( \int_{U_{j-1,j}} \mu_{a_{j-1}}(u)\mu_{a_j}(u)N(u)du + \\ &+ \int_{U_{j,j+1}} \mu_{a_j}(u)\mu_{a_{j+1}}(u)N(u)du \Big) + \\ &+ p_t \frac{1}{|U|} \int_{U_{t-1,t}} \mu_{a_{t-1}}(u)\mu_{a_t}(u)N(u)du = \\ &= \frac{1}{|U|} \sum_{j=1}^{t-1} (p_j + p_{j+1}) \int_{U_{j,j+1}} \mu_{a_j}(u)\mu_{a_{j+1}}(u)N(u)du. \end{aligned} \tag{3.16}$$

Теперь мы можем приступить к изучению основного вопроса данного раздела: связи степени нечеткости признака с объемом средних индивидуальных потерь информации и шумов.

## 3.3. Методика выбора оптимального множества значений качественных признаков

Рассмотрим частный случай $t = 2$.

**Теорема 5** *Пусть* $s_2 \in G_2(\overline{L})$; $\xi(s_t)$ *- степень нечеткости* $s_t$; $\Phi_X(U)$, $\Psi_X(U)$ *- средние индивидуальные потери информации и шумы, возникающие при поиске информации по признаку с множеством значений* $X$*, моделью которого является* $s_t$*. Пусть, далее,* $N(U)$ *- число объектов, описания которых хранятся в базе данных системы, имеющих физическое значение признака, равное* $u$*, есть константа. Тогда*

$$\Phi_X(U) = \Psi_X(U) = \frac{N}{3}\xi(s_2), N = Const.$$

**Доказательство.** Так как $s_2 \in G_2(\overline{L})$, то функции принадлежности $\mu_{a_1}(u)$ и $\mu_{a_2}(u)$ имеют вид ( A.2) и ( A.3) соответственно.

Тогда, подставляя ( A.2) и ( A.3) в ( 3.8), получаем:

$$\Phi_X(U) = \Psi_X(U) = \frac{1}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}} \frac{1}{d^2}(\overline{u}_{1R} - u)(u - \overline{u}_{2L})N(u)du. \tag{3.17}$$

По условию теоремы $N(u) = N = Const$, поэтому:

$$
\begin{aligned}
\Phi_X(U) &= \Psi_X(U) = \frac{N}{d^2|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}}(-u^2+u(\overline{u}_{2L}+\overline{u}_{1R}) \\
&\quad -\overline{u}_{1R}\overline{u}_{2L})du = \\
&= \frac{N}{d^2|U|}\Big[-\frac{u^3}{3}\Big|_{\overline{u}_{2L}}^{\overline{u}_{1R}}+(\overline{u}_{2L}+\overline{u}_{1R})\frac{u^2}{2}\Big|_{\overline{u}_{2L}}^{\overline{u}_{1R}} \\
&\quad -\overline{u}_{1R}\overline{u}_{2L}u\Big|_{\overline{u}_{2L}}^{\overline{u}_{1R}}\Big] = \\
&= \frac{N}{6d^2|U|}\Big[-2(\overline{u}_{1R}^3-\overline{u}_{2L}^3) \\
&\quad +3(\overline{u}_{2L}+\overline{u}_{1R})(\overline{u}_{1R}^2-\overline{u}_{2L}^2)- \\
&\quad -6\overline{u}_{1R}\overline{u}_{2L}(\overline{u}_{1R}-\overline{u}_{2L})\Big] = \\
&= \frac{N}{6d^2|U|}\Big[-2\overline{u}_{1R}^3+2\overline{u}_{2L}^3+3\overline{u}_{1R}^2\overline{u}_{2L}+3\overline{u}_{1R}^3- \\
&\quad -3\overline{u}_{2L}^3-3\overline{u}_{1R}\overline{u}_{2L}^2-6\overline{u}_{1R}^2\overline{u}_{2L}+6\overline{u}_{1R}\overline{u}_{2L}^2\Big] = \\
&= \frac{N(\overline{u}_{1R}-\overline{u}_{2L})^3}{6d^2|U|} = \frac{Nd^3}{6d^2|U|} = \frac{Nd}{6|U|} \qquad (3.18)
\end{aligned}
$$

Сравнивая получившееся выражение с ( A.4), получаем:

$$\Phi_X(U) = \Psi_X(U) = \frac{N}{3}\left(\frac{d}{2|U|}\right) = \frac{N}{3}\xi(s_t) \qquad (3.19)$$

Теорема полностью доказана.

Этот результат довольно просто обобщается на случай $t > 2$.

**Теорема 6** *Пусть $s_t \in G_t(\overline{L})$ и для нее выполняются условия теоремы 5. Пусть, далее, для пользователя значения признака представляют одинаковый интерес, то есть вероятности запросов по каждому значению признака равны. Тогда*

$$\Phi_X(U) = \Psi_X(U) = \frac{2N}{3t}\xi(s_t), N = Const.$$

**Доказательство.** Так как по условию теоремы $N(u) = N = Const$, ( 3.16) можно переписать следующим образом:

$$\Phi_X(U) = \Psi_X(U) = \frac{2N}{|U|} \sum_{j=1}^{t-1} \frac{1}{t} \int_{U_{j,j+1}} \mu_{a_j}(u) \mu_{a_{j+1}}(u) du. \quad (3.20)$$

Обозначим $|U_{j,j+1}| = d_{j,j+1}$, и, аналогично ( 3.18) из ( 3.20) получаем:

$$\Phi_X(U) = \Psi_X(U) = \frac{2N}{|U|} \sum_{j=1}^{t-1} \frac{1}{t} \frac{d_{j,j+1}}{6} = \frac{N}{3t|U|} \sum_{j=1}^{t-1} d_{j,j+1} = \frac{ND}{3t|U|}, \quad (3.21)$$

где $D = \sum_{j=1}^{t-1} d_{j,j+1} = \sum_{j=1}^{t-1} |U_{j,j+1}|$ - мощность подмножества в $U$, на котором $\mu_{a_j}(u) \neq 1 \forall j (1 \leq j \leq t)$. Сравнивая ( 3.16) и ( 3.21), получаем следующее обобщение ( 3.19):

$$\Phi_X(U) = \Psi_X(U) = \frac{2N}{3t} \xi(s_t). \quad (3.22)$$

Теорема полностью доказана.

Полученные в теоремах 5, 6 результаты можно обобщить для класса $G_t(L)$.

**Теорема 7** *Пусть* $s_2 \in G_2(L)$, $N(u) = N = Const$. *Тогда*

$$\Phi_X(U) = \Psi_X(U) = c\xi(s_2),$$

*где* $c$ *- константа, зависящая только от* $N$.

**Доказательство.** Так как $s_2 \in G_2(L)$, то $\forall u \in U \mu_{a_1}(u) = 1 - \mu_{a_2}(u)$. Поэтому ( 3.8) можно переписать следующим образом:

$$\Phi_X(U) = \Psi_X(U) = \frac{1}{|U|} \int_U \mu_{a_1}(u)(1 - \mu_{a_1}(u)) N(u) du \quad (3.23)$$

Рассмотрим универсум $U$ (рис. 3.2).

Ясно, что на отрезках $[\overline{u}_{1L}, \overline{u}_{2L}]$ и $[\overline{u}_{1R}, \overline{u}_{2R}]$ $\Phi_X(U) = \Psi_X(U) = \xi(s_t) = 0$. Таким образом, необходимо сравнить интегралы

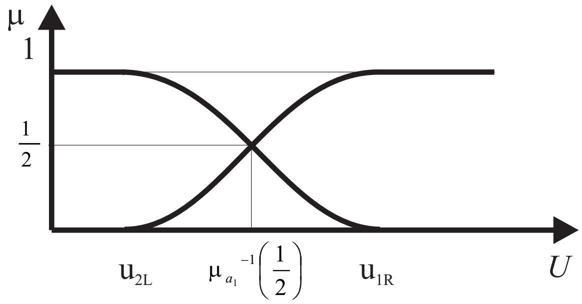


Рис. 3.2 Графическое изображение $s_2 \in G_2(L)$

$$\frac{1}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}}(1-(\mu_{i_1^*}(u)-\mu_{i_2^*}(u)))du$$

и

$$\frac{1}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}}\mu_{a_1}(u)(1-\mu_{a_1}(u))du.$$

где $\mu_{i_1^*}(u)$, $\mu_{i_2^*}(u)$ описываются формулами ( 2.5).

Область интегрирования - отрезок $[\overline{u}_{2L}, \overline{u}_{1R}]$ можно представить следующим образом:

$$[\overline{u}_{2L}, \overline{u}_{1R}] = [\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})] \cup [\mu_{a_1}^{-1}(\frac{1}{2}), \overline{u}_{1R}] \tag{3.24}$$

Рассмотрим первый отрезок правой части ( 3.24).

Так как $s_2 \in G_2(L)$, то на этом отрезке $\mu_{i_1^*}(u) = \mu_{a_1}(u)$ и $\mu_{i_2^*}(u) = \mu_{a_2}(u)$.

Обозначим через $\xi(s_2)\Big|_{[a,b]}$ значение функционала $\xi(s_2)$ на отрезке $[a,b]$ (то есть значение соответствующего интеграла для $u \in [a,b]$). Аналогичные обозначения будем использовать и для $\Phi_X(U)$, $\Psi_X(U)$. Тогда, с учетом вышесказанного, можем написать:

$$
\begin{aligned}
\xi(s_2)\Big|_{[\overline{u}_{2L},\mu_{a_1}^{-1}(\frac{1}{2})]} &= \frac{1}{|U|}\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})}(1-(\mu_{i_1^*}(u)-\mu_{i_2^*}(u)))du = \\
&= \frac{1}{|U|}\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})}(1-(\mu_{a_1}(u)-\mu_{a_2}(u)))du = \\
&= \frac{1}{|U|}\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})}(1-(\mu_{a_1}(u)-(1-\mu_{a_1}(u))))du = \\
&= \frac{2}{|U|}\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})}(1-\mu_{a_1}(u))du, \qquad (3.25)
\end{aligned}
$$

$$
\begin{aligned}
\Phi_X(U)\Big|_{[\overline{u}_{2L},\mu_{a_1}^{-1}(\frac{1}{2})]} &= \Psi_X(U)\Big|_{[\overline{u}_{2L},\mu_{a_1}^{-1}(\frac{1}{2})]} = \\
&= \frac{N}{|U|}\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})}\mu_{a_1}(u)(1-\mu_{a_1}(u))du.
\end{aligned}
\qquad (3.26)
$$

Рассмотрим разность $\Phi_X(U)\Big|_{[\overline{u}_{2L},\mu_{a_1}^{-1}(\frac{1}{2})]} - \xi(s_2)\Big|_{[\overline{u}_{2L},\mu_{a_1}^{-1}(\frac{1}{2})]}$:

$$
\begin{aligned}
\Phi_X(U)\Big|_{[\overline{u}_{2L},\mu_{a_1}^{-1}(\frac{1}{2})]} &- \xi(s_2)\Big|_{[\overline{u}_{2L},\mu_{a_1}^{-1}(\frac{1}{2})]} = \\
&= \frac{1}{|U|}\Big[N\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})}\mu_{a_1}(u)(1-\mu_{a_1}(u))du - \\
&- 2\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})}(1-\mu_{a_1}(u))du\Big]. \qquad (3.27)
\end{aligned}
$$

В математическом анализе известна первая формула среднего значения в обобщенной форме [11]. Ее суть в следующем. пусть функции $f(x)$ и $g(x)$ интегрируемы на отрезке $[a,b]$, и пусть $m$ и $M$ - точные грани $f(x)$ на $[a,b]$. Пусть, кроме того, функция $g(x) \geq 0$ (или $g(x) \leq 0$) на всем отрезке $[a,b]$. Тогда найдется такое число $\theta$, удовлетворяющее неравенствам $m \leq \theta \leq M$, что

$$\int_a^b f(x)g(x)dx = \theta \int_a^b g(x)dx \tag{3.28}$$

В частности, если $f(x)$ непрерывна на $[a, b]$, то на этом отрезку существует такое число $x^*$, что

$$\int_a^b f(x)g(x)dx = f(x^*) \int_a^b g(x)dx \tag{3.29}$$

Последнее выражение и есть первая формула среднего значения в обобщенной форме.

Так как $\mu_{a_1}(u)$ непрерывна на $[\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})]$ и $(1-\mu_{a_1}(u)) \geq 0$ на этом отрезке, то можно применить ( 3.29) к первому интегралу в правой части ( 3.27).

$$\begin{aligned}
&\Phi_X(U)\Big|_{[\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})]} - \xi(s_2)\Big|_{[\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})]} = \\
&= \frac{1}{|U|}\Big[N\mu_{a_1}(u^*) \int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})} (1-\mu_{a_1}(u))du - \\
&\qquad -2\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})} (1-\mu_{a_1}(u))du\Big] = \\
&= \frac{2}{|U|}\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})} (1-\mu_{a_1}(u))du \left[\frac{N}{2}\mu_{a_1}(u^*) - 1\right]
\end{aligned} \tag{3.30}$$

Из последнего соотношения получаем:

$$\Phi_X(U)\Big|_{[\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})]} = \frac{N}{2}\mu_{a_1}(u^*)\xi(s_2)\Big|_{[\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})]} \tag{3.31}$$

Обозначим

$$c_1 = \frac{N}{2}\mu_{a_1}(u^*), \tag{3.32}$$

где $u^* \in [\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})]$.

Тогда ( 3.31) можно переписать следующим образом:

$$\Phi_X(U)\Big|_{[\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})]} = c_1\xi(s_2)\Big|_{[\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})]}. \tag{3.33}$$

Рассмотрим второй отрезок правой части ( 3.24). Очевидно, на этом отрезке $\mu_{i_1^*}(u) = \mu_{a_2}(u)$ и $\mu_{i_2^*}(u) = \mu_{a_1}(u)$. Поэтому

$$\begin{aligned}
\xi(s_2)\Big|_{[\mu_{a_1}^{-1}(\frac{1}{2}),\overline{u}_{1R}]} &= \frac{1}{|U|}\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} (1-(\mu_{a_2}(u)-\mu_{a_1}(u)))du = \\
&= \frac{1}{|U|}\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} (1-((1-\mu_{a_1}(u))-\mu_{a_1}(u)))du = \\
&= \frac{2}{|U|}\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)du. \qquad (3.34)
\end{aligned}$$

Повторяя рассуждения для отрезка $[\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})]$, получим:

$$\begin{aligned}
&\Phi_X(U)\Big|_{[\mu_{a_1}^{-1}(\frac{1}{2}),\overline{u}_{1R}]} - \xi(s_2)\Big|_{[\mu_{a_1}^{-1}(\frac{1}{2}),\overline{u}_{1R}]} = \\
&= \frac{1}{|U|}\Big[N\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)(1-\mu_{a_1}(u))du - \\
&\qquad\qquad -2\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)du\Big] = \\
&= \frac{1}{|U|}\Big[N(1-\mu_{a_1}(u^*))\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)du \\
&\qquad\qquad -2\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)du\Big] = \\
&= \left[\frac{N}{2}(1-\mu_{a_1}(u^*))-1\right]\frac{2}{|U|}\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)du = \\
&= \left[\frac{N}{2}(1-\mu_{a_1}(u^*))-1\right]\xi(s_2)\Big|_{[\mu_{a_1}^{-1}(\frac{1}{2}),\overline{u}_{1R}]}. \qquad (3.35)
\end{aligned}$$

Таким образом,

$$\begin{aligned}
\Phi_X(U)\Big|_{[\mu_{a_1}^{-1}(\frac{1}{2}),\overline{u}_{1R}]} &= \frac{N}{2}(1-\mu_{a_1}(u^*))\xi(s_2)\Big|_{[\mu_{a_1}^{-1}(\frac{1}{2}),\overline{u}_{1R}]} = \\
&= c_2\xi(s_2)\Big|_{[\mu_{a_1}^{-1}(\frac{1}{2}),\overline{u}_{1R}]}, \qquad (3.36)
\end{aligned}$$

где

$$c_2 = \frac{N}{2}(1-\mu_{a_1}(u^*)), u^* \in [\mu_{a_1}^{-1}(\frac{1}{2}), \overline{u}_{1R}]. \qquad (3.37)$$

Обозначим через $f_1(u)$, $f_2(u)$ следующие функции, определенные на $[\overline{u}_{2L}, \overline{u}_{1R}]$ :

$$f_1(u) = \begin{cases} c_1, & u \in [\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})] \\ 0, & u \in [\mu_{a_1}^{-1}(\frac{1}{2}), \overline{u}_{1R}] \end{cases}$$

$$f_2(u) = \begin{cases} 0, & u \in [\overline{u}_{2L}, \mu_{a_1}^{-1}(\frac{1}{2})] \\ c_2, & u \in [\mu_{a_1}^{-1}(\frac{1}{2}), \overline{u}_{1R}] \end{cases}$$

Суммируя ( 3.31) и ( 3.37) с учетом ( 3.24), получаем:

$$\begin{aligned}
& \frac{N}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}} \mu_{a_1}(u)(1-\mu_{a_1}(u))du = \\
= \;& \frac{N}{|U|}\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})} \mu_{a_1}(u)(1-\mu_{a_1}(u))du \\
& + \frac{N}{|U|}\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)(1-\mu_{a_1}(u))du = \\
= \;& c_1\frac{1}{|U|}\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})} \eta(s_2,u)du + c_2\frac{1}{|U|}\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \eta(s_2,u)du = \\
= \;& \frac{1}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}} f_1(u)\eta(s_2,u)du + \frac{1}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}} f_2(u)\eta(s_2,u)du = \\
= \;& \frac{1}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}} [f_1(u)+f_2(u)]\,\eta(s_2,u)du. \qquad (3.38)
\end{aligned}$$

Применяя к правой части ( 3.38) формулу ( 3.28) (не трудно видеть, что условия применения этой формулы выполняются), получаем:

$$\frac{N}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}} \mu_{a_1}(u)(1-\mu_{a_1}(u))du = c_3\frac{1}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}} \eta(s_2,u)du. \qquad (3.39)$$

Последнее выражение можно переписать как

$$\Phi_X(U) = \Psi_X(U) = c_3\xi(s_2), \tag{3.40}$$

где $c_3$ - некоторая константа.

Из ( 3.32), ( 3.37) следует, что $\frac{N}{4} \le c_1 \le \frac{N}{2}$, $\frac{N}{4} \le c_2 \le \frac{N}{2}$, поэтому и $\frac{N}{4} \le c_3 \le \frac{N}{2}$.

Теорема доказана.

Последний результат, аналогично теореме 6, обобщается на случай $t > 2$. Действительно, справедлива следующая теорема.

**Теорема 8** *Пусть $s_t \in G_t(L)$, $N(u) = N = Const$ и вероятности запросов по каждому значению признака равны. Тогда*

$$\Phi_X(U) = \Psi_X(U) = \frac{c}{t}\xi(s_t),$$

*где $c$ - некоторая константа, зависящая только от $N$.*

**Доказательство.** Для сформулированных условий справедлива формула ( 3.20) для $\Phi_X(U)$ и $\Psi_X(U)$:

$$\Phi_X(U) = \Psi_X(U) = \frac{2N}{|U|}\sum_{j=1}^{t-1}\int_{U_{j,j+1}} \mu_{a_j}(u)\mu_{a_{j+1}}(u)du$$

Используя разбиение ( refproof: 317), можно написать аналогичное выражение и для $\xi(s_t)$:

$$\xi(s_t) = \frac{1}{|U|}\sum_{j=1}^{t-1}\int_{U_{j,j+1}} \eta(s_t, u)du.$$

Зафиксируем некоторый номер $j(1 \le j \le t-1)$. На основании теоремы ( 7), справедливо следующее равенство:

$$\frac{N}{|U|}\int_{U_{j,j+1}} \mu_{a_j}(u)\mu_{a_{j+1}}(u)du = c_j\frac{1}{|U|}\int_{U_{j,j+1}} \eta(s_t, u)du,$$

где $c_j$ - некоторая константа, зависящая только от $N$.

Таким образом,

$$\sum_{j=1}^{t-1}\frac{N}{|U|}\int_{U_{j,j+1}}\mu_{a_j}(u)\mu_{a_{j+1}}(u)du=\sum_{j=1}^{t-1}c_j\frac{1}{|U|}\int_{U_{j,j+1}}\eta(s_t,u)du$$

Умножая последнее равенство на $\frac{2}{t}$, получим:

$$\frac{N}{|U|}\frac{2}{t}\sum_{j=1}^{t-1}\int_{U_{j,j+1}}\mu_{a_j}(u)\mu_{a_{j+1}}(u)du=\frac{2}{t|U|}\sum_{j=1}^{t-1}c_j\int_{U_{j,j+1}}\eta(s_t,u)du.$$

Обозначим через $f(u)$ следующую функцию, определенную на $U$:

$$f(u)=\begin{cases} c_j, & u\in U_{j,j+1}(1\le j\le t-1)\\ 0, & u\in U_j(1\le j\le t)\end{cases}$$

Тогда последнюю формулу можно переписать следующим образом:

$$\Phi_X(U)=\Psi_X(U)=\frac{2}{t|U|}\int_U f(u)\eta(s_t,u)du$$

Применяя к правой части последней формулы ( 3.28) (не трудно видеть, что условия применимости этой формулы выполняются), получаем:

$$\Phi_X(U)=\Psi_X(U)=\frac{2}{t|U|}c\int_U\eta(s_t,u)du=\frac{2c}{t}\xi(s_t).$$

Так как $\frac{N}{4}\le c_j\le\frac{N}{2}$, то и $\frac{N}{4}\le c\le\frac{N}{2}$.

Теорема доказана.

Таким образом, уменьшение степени нечеткости на $\alpha$ процентов при фиксированном числе значений поисковой характеристики приводит к такому же сокращению средних индивидуальных потерь информации и шумов. Уменьшение же степени нечеткости при одновременном увеличении числа значений характеристики делает эту зависимость еще более сильной.

На основе теорем данного раздела можно предложить следующую методику выбора множества значений качественного признака.

1. Формируются все возможные множества значений признака.

2. Каждое множество значений признака представляется в виде полного ортогонального семантического пространства.

3. Для каждого множества значений вычисляется степень нечеткости ПОСП.

4. В качестве оптимального множества значений, обеспечивающего максимальное качество поиска информации по этому признаку, выбирается то множество, для которого отношение степени нечеткости к числу значений признака минимально.

При поиске информации по совокупности признаков необходимо обобщить полученные результаты. Можно показать [1)] , что в случае двух поисковых характеристик $X^1$ и $X^2$, содержащих $t_1$ и $t_2$ значений соответственно,

$$\begin{aligned} \Phi_{X^1,X^2}(U^1,U^2) &= \frac{1}{t_1}\Phi_{X^2}(U^2)+\frac{1}{t_2}\Phi_{X^1}(U^1)- \\ &- \frac{1}{N}\Phi_{X^1}(U^1)\Phi_{X^2}(U^2). \end{aligned} \tag{3.41}$$

Применяя к правой части данного равенства теорему ( 8), получим, что при выполнении условий этой теоремы

$$\begin{aligned} \Phi_{X^1,X^2}(U^1,U^2) &= \frac{c_1}{t_1t_2}\xi(s^2_{t_2})+\frac{c_2}{t_1t_2}\xi(s^1_{t_1})- \\ &- \frac{c_1c_2}{t_1t_2N}\xi(s^1_{t_1})\xi(s^2_{t_2}), \end{aligned} \tag{3.42}$$

где $c_1$, $c_2$ - некоторые константы, зависящие только от $N$, $s^1_{t_1}$ и $s^2_{t_2}$ - представление множеств значений $X^1$ и $X^2$ в виде ПОСП.

Таким образом, чтобы минимизировать потери информации и шумы, возникающие при поиске информации по двум заданным признакам, необходимо так выбирать множества значений признаков $X^1$ и $X^2$, чтобы

$$\frac{1}{c_1c_2}\left[c_1\xi(s^2_{t_2})+c_2\xi(s^1_{t_1})-c_3\xi(s^1_{t_1})\xi(s^2_{t_2})\right]\longrightarrow \min .$$

1) Доказательство данного равенства не приводится в силу его громоздкости.

Вполне очевидно, что могут существовать конкретные ситуации, когда множества значений признаков $X^1$ и $X^2$, оптимальные с точки зрения последнего критерия (то есть совместного поиска) могут не совпадать с множествами значений, оптимальных с точки зрения поиска по каждому значению признака в отдельности. Окончательный выбор множеств значений зависит от конкретной задачи (в частности, от вероятности запросов по первому и второму признаку, по их совокупности и т.п.).

### 3.3.1. Устойчивость методики выбора оптимального множества значений

Одним из ограничений методики выбора оптимального множества значений качественных признаков (раздел 3.3), существенно используемых при ее анализе, является предположение об одинаковости функций принадлежности используемых лингвистических понятий у источника информации и пользователя системы. Интуитивно ясно, что функции принадлежности у всех людей не могут быть полностью одинаковыми. Данный тезис можно проиллюстрировать следующими примерами.

Если попросить оценить возраст юношу и пожилого человека, то, скорее всего, их оценки будут различаться: для молодого человека 50 лет - это старость, для пожилого - зрелый возраст. То же самое может наблюдаться при оценке роста низким и высоким человеком: функции принадлежности соответствующих лингвистических понятий у первого будут сдвинуты влево, у второго - вправо относительно друг друга. Подобные ситуации давно были замечены и даже послужили основой сюжета ряда детективов.

Более сложное взаимодействие между семантикой одинаковых терминов может наблюдаться у людей разных национальностей (проживающих в различных геоклиматических зонах) в силу принципа лингвистической дополнительности - одного из принципов психолингвистики [5]. Так, например, жители крайнего севера (эскимосы, чукчи) различают несколько оттенков снега, соответствующих его состоянию, финны имеют несколько названий для синего цвета.

Для отражения перечисленных факторов будем считать, что

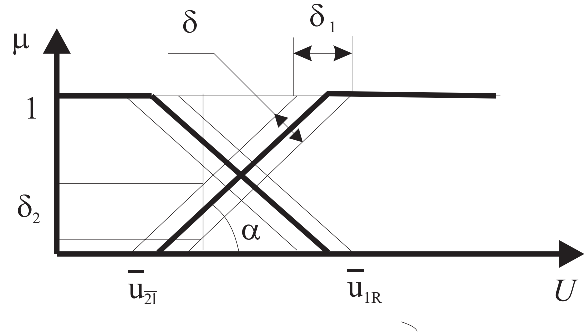


Рис. 3.3 Графическое изображение функций принадлежности в $\delta-$ модели.

функции принадлежностей источника информации и пользователя не совпадают, а могут находиться в некоторой полосе ширины $\delta$, то есть заданы с некоторой "точностью" $\delta$ (рис. 3.3).

Выразим основные параметры ($\delta_1$ и $\delta_2$), необходимые для анализа модели, как функции от $\delta$. Для этого воспользуемся элементарными соотношениями из тригонометрии.

Обозначим через $\alpha$ угол наклона $\mu_{a_2}(u)$ в точке $\overline{u}_{2L}$ (рис. 3.3). Тогда $\tan\alpha = \frac{1}{d}$, где $d = \overline{u}_{1R} - \overline{u}_{2L}$. Рассмотрим более подробно треугольник $ABC$, где $\angle BAC = \alpha$, $|BC| = \frac{\delta_2}{2}$, $|CD| = \frac{\delta}{2}$, $|AC| = \frac{\delta_1}{2}$ (рис. 3.4).

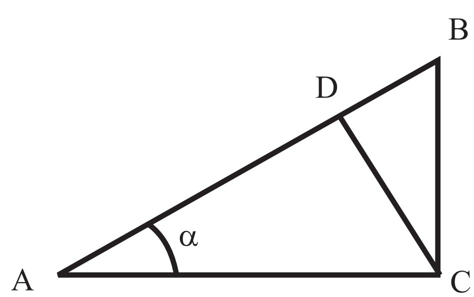


Рис. 3.4 Анализ $\delta-$ модели.

Из $\triangle ABC$ следует, что $\tan\alpha = \frac{|BC|}{|AC|} = \frac{\delta_2}{\delta_1}$.

Таким образом, $\frac{\delta_2}{\delta_1} = \frac{1}{d}$, и, следовательно,

$$\delta_1 = \delta_2 d. \tag{3.43}$$

Выразим $\delta_2$ через $\delta$ и $d$. С помощью простейших тригонометрических соотношений из $\triangle CDB$ и $\triangle ADC$, получаем, что $\delta^2(1+d^2) = \delta_2^2 d^2$.

Из последнего равенства получаем:

$$\delta_2^2 = \frac{\delta^2}{d^2}(1+d^2)$$

или

$$\delta_2 = \frac{\delta}{d}\sqrt{1+d^2}. \tag{3.44}$$

Вспоминая соотношение $\delta_1 = \delta_2 d$, получаем:

$$\delta_1 = \delta\sqrt{1+d^2}. \tag{3.45}$$

Имея выражения для $\delta_2$ и $\delta_1$ (( 3.44) и ( 3.45) соответственно) можно оценить средние индивидуальные потери информации и шумы для описанной модели (будем называть ее $\delta$ - модель), а также степень нечеткости соотвнтствующего ПОСП.

### Потери информации и шумы в $\delta$ - модели

Рассмотрим аналогично разделу 3.2 простейший случай $t = 2$ (рис. 3.5).

Рассмотрим, так же как и в разделе 3.2 некоторую точку

$$u_1^\star \in \left[\overline{u}_{2L} + \frac{\delta_1}{2}, \overline{u}_{1R} - \frac{\delta_1}{2}\right]$$

и запрос $\langle I(O) = a_1\rangle$. В этом случае

$$\begin{aligned}
\mu_{a_1}^L(u_1^\star)N(u_1^\star) &\le N_{a_1}(u_1^\star) \le \mu_{a_1}^R(u_1^\star)N(u_1^\star)\\
\mu_{a_2}^R(u_1^\star)N(u_1^\star) &\le N_{a_2}(u_1^\star) \le \mu_{a_2}^L(u_1^\star)N(u_1^\star)\\
\mu_{a_1}^L(u_1^\star)N^E &\le N_{a_1}^E(u_1^\star) \le \mu_{a_1}^R(u_1^\star)N^E\\
\mu_{a_2}^R(u_1^\star)N^E &\le N_{a_2}^E(u_1^\star) \le \mu_{a_2}^L(u_1^\star)N^E.
\end{aligned} \tag{3.46}$$

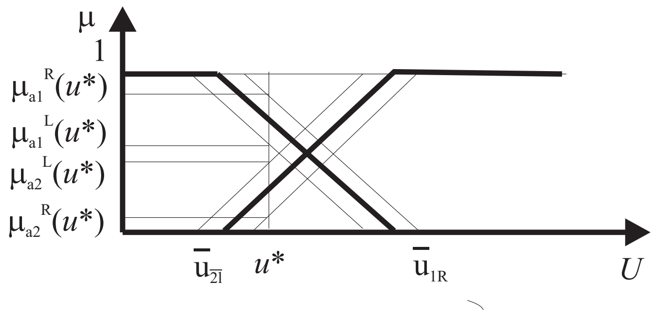


Рис. 3.5 Анализ $\delta-$ модели.

В ( 3.46) через $N^E$, $N(u_1^\star)$ аналогично 3.2 обозначены число пользователей и число объектов в базе данных системы, имеющих физическое значение признака, равное $u_1^\star$; $\mu_{a_i}^L(u_1^\star)$, $\mu_{a_i}^R(u_1^\star)$ $(i=1,2)$ - левая и правая граница функций принадлежности $\mu_{a_i}(u)$ (рис. 3.5).

Повторив рассуждения раздела 3.2, получаем:

$$\begin{cases} \varphi_{a_1}(u_1^\star) \geq \frac{1}{N^E}\mu_{a_1}^L(u_1^\star)N^E\mu_{a_2}^R(u_1^\star)N(u_1^\star) \\ \varphi_{a_1}(u_1^\star) \leq \frac{1}{N^E}\mu_{a_1}^R(u_1^\star)N^E\mu_{a_2}^L(u_1^\star)N(u_1^\star) \end{cases} \tag{3.47}$$

$$\begin{cases} \psi_{a_1}(u_1^\star) \geq \frac{1}{N^E}\mu_{a_2}^R(u_1^\star)N^E\mu_{a_1}^L(u_1^\star)N(u_1^\star) \\ \psi_{a_1}(u_1^\star) \leq \frac{1}{N^E}\mu_{a_2}^L(u_1^\star)N^E\mu_{a_1}^R(u_1^\star)N(u_1^\star) \end{cases} \tag{3.48}$$

Рассмотрим точку

$$u_2^\star \in \left[\overline{u}_{2L} - \frac{\delta_1}{2}, \overline{u}_{2L} + \frac{\delta_1}{2}\right].$$

На этом отрезке $\mu_{a_1}^R(u_2^\star) = 1$, $\mu_{a_2}^R(u_2^\star) = 0$, поэтому из ( 3.47), ( 3.48) непосредственно получаем:

$$0 \leq \varphi_{a_1}(u_2^\star) \leq \mu_{a_2}^L(u_2^\star)N(u_2^\star), \tag{3.49}$$

$$0 \le \psi_{a_1}(u_2^\star) \le \mu_{a_2}^L(u_2^\star) N(u_2^\star). \tag{3.50}$$

Аналогично для

$$u_3^\star \in \left[\overline{u}_{1R} - \frac{\delta_1}{2}, \overline{u}_{1R} + \frac{\delta_1}{2}\right]$$

$(\mu_{a_1}^L(u_3^\star) = 0,\ \mu_{a_2}^L(u_3^\star) = 1)$

$$0 \le \varphi_{a_1}(u_3^\star) \le \mu_{a_1}^R(u_3^\star) N(u_3^\star), \tag{3.51}$$

$$0 \le \psi_{a_1}(u_3^\star) \le \mu_{a_1}^R(u_3^\star) N(u_3^\star). \tag{3.52}$$

Повторив рассуждения вывода формулы ( 3.3), мы сможем построить нижнюю ($\underline{\Phi}_{a_1}(U)$) и верхнюю ($\overline{\Phi}_{a_1}(U)$) оценки $\Phi_{a_1}(U)$:

$$\underline{\Phi}_{a_1}(U) = \frac{1}{|U|} \int_{\overline{u}_{2L} + \frac{\delta_1}{2}}^{\overline{u}_{1R} - \frac{\delta_1}{2}} \mu_{a_1}^L(u) \mu_{a_2}^R(u) N(u) du, \tag{3.53}$$

$$\begin{aligned} \overline{\Phi}_{a_1}(U) &= \frac{1}{|U|} \Big[ \int_{\overline{u}_{2L} - \frac{\delta_1}{2}}^{\overline{u}_{2L} + \frac{\delta_1}{2}} \mu_{a_2}^L(u) N(u) du + \\ &+ \int_{\overline{u}_{2L} + \frac{\delta_1}{2}}^{\overline{u}_{1R} - \frac{\delta_1}{2}} \mu_{a_1}^R(u) \mu_{a_2}^L(u) N(u) du + \int_{\overline{u}_{1R} - \frac{\delta_1}{2}}^{\overline{u}_{1R} + \frac{\delta_1}{2}} \mu_{a_1}^R(u) N(u) \Big]. \end{aligned} \tag{3.54}$$

Аналогично для $\Psi_{a_1}(U)$

$$\underline{\Psi}_{a_1}(U) = \frac{1}{|U|} \int_{\overline{u}_{2L} + \frac{\delta_1}{2}}^{\overline{u}_{1R} - \frac{\delta_1}{2}} \mu_{a_1}^L(u) \mu_{a_2}^R(u) N(u) du, \tag{3.55}$$

$$
\begin{aligned}
\overline{\Psi}_{a_1}(U) &= \frac{1}{|U|}\Big[\int_{\overline{u}_{2L}-\frac{\delta_1}{2}}^{\overline{u}_{2L}+\frac{\delta_1}{2}} \mu^L_{a_2}(u)N(u)du + \\
&+ \int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} \mu^L_{a_2}(u)\mu^R_{a_1}(u)N(u)du + \int_{\overline{u}_{1R}-\frac{\delta_1}{2}}^{\overline{u}_{1R}+\frac{\delta_1}{2}} \mu^R_{a_1}(u)N(u)\Big].
\end{aligned}
\tag{3.56}
$$

Сравнивая ( 3.53) и ( 3.55) получаем, что $\underline{\Phi}_{a_1}(U) = \underline{\Psi}_{a_1}(U)$; ( 3.54) и ( 3.56) - что $\overline{\Phi}_{a_1}(U) = \overline{\Psi}_{a_1}(U)$.

Вычислим верхние и нижние оценки $\underline{\Phi}_{a_1}(U)$ и $\overline{\Phi}_{a_1}(U)$ для $N(u) = N = Const$.

$$
\begin{aligned}
\underline{\Phi}_{a_1}(U) &= \frac{1}{|U|}\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} \frac{1}{d}\left(u-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\right)\times \\
&\qquad \times\frac{1}{d}\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)-u\right)N du = \\
&= \frac{N}{d^2|U|}\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} \left(u\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)-u^2+u\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)-\right. \\
&\qquad \left.-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)\right)du = \\
&= \frac{N}{d^2|U|}\left[-\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} u^2 du+\left(\overline{u}_{1R}-\frac{\delta_1}{2}+\overline{u}_{2L}+\frac{\delta_1}{2}\right)\times\right. \\
&\quad \left.\times\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} u du-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} du\right] = \\
&= \frac{N}{d^2|U|}\left[-\frac{1}{3}\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)^3-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)^3\right)+\right. \\
&\qquad +\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)+\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\right)\times \\
&\quad \times\frac{1}{2}\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)^2-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)^2\right)- \\
&\left.-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\right)\right]= \\
&= \frac{N}{6d^2|U|}\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\right)^3=\frac{N\left(d-\delta_1\right)^3}{6d^2|U|}
\end{aligned}
\tag{3.57}
$$

С учетом ( 3.44), ( 3.45) из ( 3.57) непосредственно получаем:

$$\begin{aligned}
\underline{\Phi}_{a_1}(U) &= \frac{N\left(d-\delta\sqrt{1+d^2}\right)^3}{6d^2|U|} = \frac{Nd\left(1-\frac{\delta}{d}\sqrt{1+d^2}\right)^3}{6|U|} = \\
&= \frac{Nd\left(1-\delta_2\right)^3}{6|U|} \qquad (3.58)
\end{aligned}$$

$$\begin{aligned}
\overline{\Phi}_{a_1}(U) &= \frac{1}{|U|}\left[\int_{\overline{u}_{2L}-\frac{\delta_1}{2}}^{\overline{u}_{2L}+\frac{\delta_1}{2}} \frac{1}{d}\left(u-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\right)Ndu+\right. \\
&+\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} \frac{1}{d}\left(u-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\right)\frac{1}{d}\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)-u\right)Ndu+ \\
&+ \left.\int_{\overline{u}_{1R}-\frac{\delta_1}{2}}^{\overline{u}_{1R}+\frac{\delta_1}{2}} \frac{1}{d}\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)-u\right)Ndu\right]= \\
&= \frac{N}{d|U|}\int_0^{\delta_1} zdz + \frac{N}{d^2|U|}\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}}\left(u\left(\left(\overline{u}_{1R}+\frac{\delta_1}{2}\right)+\right.\right. \\
&+\left.\left(\overline{u}_{2L}-\frac{\delta_1}{2}\right)\right)-u^2-\left(\overline{u}_{2L}-\frac{\delta_1}{2}\right)\left.\left(\overline{u}_{1R}+\frac{\delta_1}{2}\right)\right)du+ \\
&+\frac{N}{d|U|}\int_0^{\delta_1} ydy = \frac{2N}{d|U|}\frac{1}{2}\delta_1^2 - \frac{N}{d^2|U|}\left[\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} u^2du-\right. \\
&-\left(\left(\overline{u}_{1R}+\frac{\delta_1}{2}\right)+\left(\overline{u}_{2L}-\frac{\delta_1}{2}\right)\right)\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} udu+ \\
&+\left.\left(\overline{u}_{2L}-\frac{\delta_1}{2}\right)\left(\overline{u}_{1R}+\frac{\delta_1}{2}\right)\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} du\right]= \\
&= \frac{N\delta_1^2}{d|U|}-\frac{N}{d^2|U|}\left[\frac{1}{3}\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)^3-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)^3\right)-\right. \\
&-\left(\left(\overline{u}_{1R}+\frac{\delta_1}{2}\right)+\left(\overline{u}_{2L}-\frac{\delta_1}{2}\right)\right)\times \\
&\times\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)^2-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)^2\right)+
\end{aligned}$$

$$
\begin{aligned}
&+\left(\overline{u}_{2L}-\frac{\delta_1}{2}\right)\left(\overline{u}_{1R}+\frac{\delta_1}{2}\right)\times\\
&\qquad\times\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\right)\Bigg]=\\
=\;&\frac{N\delta_1^2}{d|U|}-\frac{N}{6d^2|U|}\left[2\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)^3-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)^3\right)-\right.\\
&-3\left(\left(\overline{u}_{1R}+\frac{\delta_1}{2}\right)+\left(\overline{u}_{2L}-\frac{\delta_1}{2}\right)\right)\times\\
&\qquad\times\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)^2-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)^2\right)+\\
&+6\left(\overline{u}_{2L}-\frac{\delta_1}{2}\right)\left(\overline{u}_{1R}+\frac{\delta_1}{2}\right)\times\\
&\qquad\left.\times\left(\left(\overline{u}_{1R}-\frac{\delta_1}{2}\right)-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\right)\right]=\\
=\;&\frac{N\delta_1^2}{d|U|}-\frac{N}{6d^2|U|}\Big[2\left(s^3-t^3\right)-3\left(s+t\right)\left(s^2-t^2\right)+\\
&\qquad+6\left(s+\delta_1\right)\left(t-\delta_1\right)\left(s-t\right)\Big]=\\
=\;&\frac{N\delta_1^2}{d|U|}-\frac{N}{6d^2|U|}\Big[2\left(s^3-t^3\right)-3\left(s+t\right)\left(s^2-t^2\right)+\\
&+6st\left(s-t\right)+6\left(s-t\right)\left(t\delta_1-s\delta_1-\delta_1^2\right)\Big]=\\
=\;&\frac{N\delta_1^2}{d|U|}-\frac{N}{6d^2|U|}\left[\left(t-s\right)^3-6\left(t-s\right)\left(t-s-\delta_1\right)\right]=\\
=\;&\frac{N\delta_1^2}{d|U|}-\frac{N}{6d^2|U|}\left[\left(\overline{u}_{2L}+\frac{\delta_1}{2}-\overline{u}_{1R}+\frac{\delta_1}{2}\right)^3-\right.\\
&-6\delta_1\left(\overline{u}_{2L}+\frac{\delta_1}{2}-\overline{u}_{1R}+\frac{\delta_1}{2}\right)\times\\
&\qquad\left.\times\left(\overline{u}_{2L}+\frac{\delta_1}{2}-\overline{u}_{1R}+\frac{\delta_1}{2}-\delta_1\right)\right]=
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{N\delta_1^2}{d|U|} - \frac{N}{6d^2|U|}\left[(\delta_1 - d)^3 + 6\delta_1(\delta_1 - d)\,d\right] = \\
&= \frac{N\delta_1^2}{d|U|} + \frac{N(d-\delta_1)^3}{6d^2|U|} + \frac{6Nd\delta_1(d-\delta_1)}{6d^2|U|} = \\
&= \frac{N}{6d^2|U|}\left((d-\delta_1)^3 + 6d\delta_1(d-\delta_1)\right). \qquad (3.59)
\end{aligned}
$$

При выводе ( 3.59) использовались замены переменных

$$z = u - \left(\overline{u}_{2L} - \frac{\delta_1}{2}\right), y = \left(\overline{u}_{1R} + \frac{\delta_1}{2}\right) - u$$

и обозначения

$$s = \overline{u}_{1R} - \frac{\delta_1}{2}, t = \overline{u}_{2L} + \frac{\delta_1}{2}.$$

Вспоминая соотношение $\delta_1 = d\delta_2$, ( 3.59) можно переписать следующим образом:

$$
\begin{aligned}
\overline{\Phi}_{a_1}(U) &= \frac{N}{6d^2|U|}\left(d^3(1-\delta_2)^3 + 6d^3\delta_2^2 + 6d^3(1-\delta_2)\,\delta_2\right) = \\
&= \frac{Nd}{6|U|}\left((1-\delta_2)^3 + 6\delta_2^2 + 6(1-\delta_2)\,\delta_2\right). \qquad (3.60)
\end{aligned}
$$

Из ( 3.58) следует, что

$$\overline{\Phi}_{a_1}(U) = \underline{\Phi}_{a_1}(U) + \frac{Nd\delta_2}{|U|}(\delta_2 + 1 - \delta_2) = \underline{\Phi}_{a_1}(U) + \frac{Nd\delta_2}{|U|}. \quad (3.61)$$

Анализируя запрос $\langle I(O) = a_2 \rangle$ из ( 3.46) получим:

$$
\begin{cases}
\varphi_{a_2}(u_1^\star) \geq \frac{1}{N^E}\mu_{a_2}^R(u_1^\star)N^E\mu_{a_1}^L(u_1^\star)N(u_1^\star) \\
\varphi_{a_2}(u_1^\star) \leq \frac{1}{N^E}\mu_{a_2}^L(u_1^\star)N^E\mu_{a_1}^R(u_1^\star)N(u_1^\star),
\end{cases} \qquad (3.62)
$$

$$
\begin{cases} \psi_{a_2}(u_1^\star) \geq \frac{1}{N^E}\mu_{a_1}^L(u_1^\star)N^E\mu_{a_2}^R(u_1^\star)N(u_1^\star) \\ \psi_{a_2}(u_1^\star) \leq \frac{1}{N^E}\mu_{a_1}^R(u_1^\star)N^E\mu_{a_2}^L(u_1^\star)N(u_1^\star). \end{cases} \tag{3.63}
$$

Для

$$
u_2^\star \in \left[\overline{u}_{2L} - \frac{\delta_1}{2}, \overline{u}_{2L} + \frac{\delta_1}{2}\right]
$$

получим следующие аналоги ( 3.49), ( 3.50):

$$
0 \leq \varphi_{a_2}(u_2^\star) \leq \mu_{a_2}^L(u_2^\star)N(u_2^\star), \tag{3.64}
$$

$$
0 \leq \psi_{a_2}(u_2^\star) \leq \mu_{a_2}^L(u_2^\star)N(u_2^\star). \tag{3.65}
$$

Аналогично для

$$
u_3^\star \in \left[\overline{u}_{1R} - \frac{\delta_1}{2}, \overline{u}_{1R} + \frac{\delta_1}{2}\right]
$$

$$
0 \leq \varphi_{a_2}(u_3^\star) \leq \mu_{a_1}^R(u_3^\star)N(u_3^\star), \tag{3.66}
$$

$$
0 \leq \psi_{a_2}(u_3^\star) \leq \mu_{a_1}^R(u_3^\star)N(u_3^\star). \tag{3.67}
$$

Учитывая полную аналогичность формул ( 3.47) - ( 3.52) и ( 3.62) - ( 3.67) получаем, что $\overline{\Phi}_{a_2}(U)$, $\underline{\Phi}_{a_2}(U)$, $\overline{\Psi}_{a_2}(U)$, $\underline{\Psi}_{a_2}(U)$ равны правым частям формул ( 3.53) - ( 3.56) соответственно.

Аналогично ( 3.6) и ( 3.7), оценки потерь информации и шумов при поиске по данному признаку будут равны:

$$
\underline{\Phi}(U) = p_1\underline{\Phi}_{a_1}(U) + p_2\underline{\Phi}_{a_2}(U),
$$

$$
\overline{\Phi}(U) = p_1\overline{\Phi}_{a_1}(U) + p_2\overline{\Phi}_{a_2}(U),
$$

$$
\underline{\Psi}(U) = p_1\underline{\Psi}_{a_1}(U) + p_2\underline{\Psi}_{a_2}(U),
$$

$$
\overline{\Psi}(U) = p_1\overline{\Psi}_{a_1}(U) + p_2\overline{\Psi}_{a_2}(U),
$$

где $p_1$ ($p_2$) - вероятность запроса по первому (второму) значению признака. Так как $p_1+p_2=1$, то из последних соотношений и равенств

$$\underline{\Phi}_{a_1}(U)=\underline{\Phi}_{a_2}(U)=\underline{\Psi}_{a_1}(U)=\underline{\Psi}_{a_2}(U)$$

и

$$\overline{\Phi}_{a_1}(U)=\overline{\Phi}_{a_2}(U)=\overline{\Psi}_{a_1}(U)=\overline{\Psi}_{a_2}(U)$$

получим, что верхние оценки средних индивидуальных потерь информации и шумов равны правой части ( 3.53), а их нижние оценки - правой части ( 3.53).

Имея результаты вычислений ( 3.53), ( 3.54), можно сформулировать и считать доказанными следующие теоремы.

**Теорема 9** *Пусть у нас есть один поисковый признак, имеющий два значения, и функции принадлежности источника информации и пользователя в каждой точке универсума отличаются не больше, чем на $\delta_2$. Тогда при $N(u)=N=Const$ нижние оценки средних индивидуальных потерь информации и шумов, возникающие при поиске информации по данному признаку*

$$\underline{\Phi}(U)=\underline{\Psi}(U)=\frac{Nd\,(1-\delta_2)^3}{6|U|}.$$

**Теорема 10** *Пусть выполняются условия теоремы 9. Тогда верхние оценки средних индивидуальных потерь информации и шумов, возникающие при поиске информации по данному признаку*

$$\overline{\Phi}(U)=\overline{\Psi}(U)=\frac{Nd\,(1-\delta_2)^3}{6|U|}+\frac{Nd\delta_2}{|U|}.$$

**Следствие 1** *Пусть выполняются условия теоремы 9. Тогда*

$$\overline{\Phi}(U)=\underline{\Phi}(U)+\frac{Nd\delta_2}{|U|},$$

$$\overline{\Psi}(U)=\underline{\Psi}(U)+\frac{Nd\delta_2}{|U|}.$$

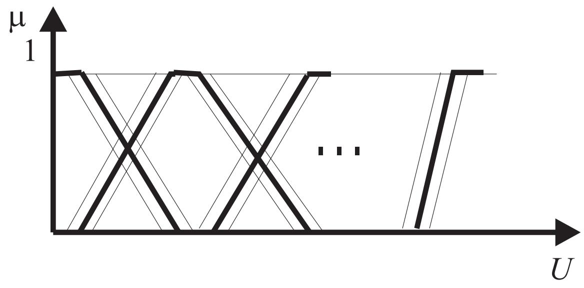


Рис. 3.6 Анализ $\delta-$ модели.

Данные результаты довольно легко обобщаются на случай более двух значений поисковой характеристики.

**Теорема 11** *Пусть у нас есть один поисковый признак, имеющий $t$ значений, и функции принадлежности источника информации и пользователя в каждой точке универсума отличаются не больше, чем на $\delta_2$. Тогда при $N(u) = N = Const$ и равной вероятности запросов по каждому значению признака нижние оценки средних индивидуальных потерь информации и шумов, возникающие при поиске информации по данному признаку*

$$\underline{\Phi}(U) = \underline{\Psi}(U) = \frac{ND\,(1-\delta_2)^3}{6|U|},$$

*где* $D = \sum_{j=1}^{t-1} d_{j,j+1}$.

**Доказательство.** Аналогично доказательству теоремы ( 6), представим универсум $U$ следующим образом (рис. 3.6)

$$
\begin{aligned}
U \quad = \quad & [\overline{u}_{1L}, \overline{u}_{t,R}] = \left[\overline{u}_{1L}, \overline{u}_{2L} - \frac{\delta_1}{2}\right] \cup \left[\overline{u}_{2L} - \frac{\delta_1}{2}, \overline{u}_{2L} + \frac{\delta_1}{2}\right] \cup \\
\cup \quad & \left[\overline{u}_{2L} + \frac{\delta_1}{2}, \overline{u}_{1R} - \frac{\delta_1}{2}\right] \cup \left[\overline{u}_{1R} - \frac{\delta_1}{2}, \overline{u}_{1R} + \frac{\delta_1}{2}\right] \cup ... \cup \\
\cup \quad & \left[\overline{u}_{jL} - \frac{\delta_1}{2}, \overline{u}_{jL} + \frac{\delta_1}{2}\right] \cup \left[\overline{u}_{jL} + \frac{\delta_1}{2}, \overline{u}_{j-1,R} - \frac{\delta_1}{2}\right] \cup \\
\cup \quad & \left[\overline{u}_{j-1,R} - \frac{\delta_1}{2}, \overline{u}_{j-1,R} + \frac{\delta_1}{2}\right] \cup \left[\overline{u}_{j-1,R} + \frac{\delta_1}{2}, \overline{u}_{j+1,L} - \frac{\delta_1}{2}\right] \cup \\
\cup \quad & \left[\overline{u}_{j+1,L} - \frac{\delta_1}{2}, \overline{u}_{j+1,L} + \frac{\delta_1}{2}\right] \cup \left[\overline{u}_{j+1,L} + \frac{\delta_1}{2}, \overline{u}_{jR} - \frac{\delta_1}{2}\right] \cup \\
\cup \quad & \left[\overline{u}_{jR} - \frac{\delta_1}{2}, \overline{u}_{jR} + \frac{\delta_1}{2}\right] \cup ... \cup \left[\overline{u}_{tL} - \frac{\delta_1}{2}, \overline{u}_{tL} + \frac{\delta_1}{2}\right] \cup \\
\cup \quad & \left[\overline{u}_{tL} + \frac{\delta_1}{2}, \overline{u}_{t-1,R} - \frac{\delta_1}{2}\right] \cup \left[\overline{u}_{t-1,R} - \frac{\delta_1}{2}, \overline{u}_{t-1,R} + \frac{\delta_1}{2}\right] \cup \\
\cup \quad & \left[\overline{u}_{t-1,R} + \frac{\delta_1}{2}, \overline{u}_{tR}\right]. \qquad (3.68)
\end{aligned}
$$

Зафиксируем некоторый номер $j(1 \le j \le t)$ и построим оценки $\underline{\Phi}_{a_j}(U)$, $\underline{\Psi}_{a_j}(U)$.

Аналогично ( 3.10), ( 3.11) представим $\underline{\Phi}_{a_j}(U)$, $\underline{\Psi}_{a_j}(U)$ в следующем виде:

$$
\underline{\Phi}_{a_j}(U) = \underline{\Phi}_{a_j}^{a_{j-1}}(U) + \underline{\Phi}_{a_j}^{a_{j+1}}(U),
$$

$$
\underline{\Psi}_{a_j}(U) = \underline{\Psi}_{a_j}^{a_{j-1}}(U) + \underline{\Psi}_{a_j}^{a_{j+1}}(U).
$$

Опираясь на доказательство теоремы ( 9) не трудно показать, что

$$
\underline{\Phi}_{a_j}^{a_{j-1}}(U) = \frac{1}{|U|} \int_{\overline{u}_{jL}+\frac{\delta_1}{2}}^{\overline{u}_{j-1,R}-\frac{\delta_1}{2}} \mu_{a_j}^R(u)\mu_{a_{j-1}}^L(u)N(u)du,
$$

$$
\underline{\Phi}_{a_j}^{a_{j+1}}(U) = \frac{1}{|U|} \int_{\overline{u}_{jL}+\frac{\delta_1}{2}}^{\overline{u}_{j+1,R}-\frac{\delta_1}{2}} \mu_{a_{j+1}}^R(u)\mu_{a_j}^L(u)N(u)du,
$$

$$\underline{\Psi}_{a_j}^{a_{j-1}}(U) = \underline{\Phi}_{a_j}^{a_{j-1}}(U),$$

$$\underline{\Psi}_{a_j}^{a_{j+1}}(U) = \underline{\Phi}_{a_j}^{a_{j+1}}(U).$$

Таким образом, при условии равной вероятности запросов

$$\begin{aligned}
\underline{\Phi}(U) &= \underline{\Psi}(U) = \\
&= \frac{1}{t}\left(\underline{\Phi}_{a_1}^{a_2}(U) + \sum_{j=2}^{t-1}\left(\underline{\Phi}_{a_j}^{a_{j-1}}(U) + \underline{\Phi}_{a_j}^{a_{j+1}}(U)\right) + \underline{\Phi}_{a_t}^{a_{t-1}}(U)\right) = \\
&= \frac{2}{t}\sum_{j=1}^{t-1}\frac{Nd_{j,j+1}\left(1-\delta_2\right)^3}{6|U|} = \frac{ND\left(1-\delta_2\right)^3}{3t|U|}, \qquad (3.69)
\end{aligned}$$

где $D = \sum_{j=1}^{t-1} d_{j,j+1}$.

Теорема доказана.

**Теорема 12** *Пусть выполняются условия теоремы 11. Тогда верхние оценки средних индивидуальных потерь информации и шумов, возникающие при поиске информации по данному признаку*

$$\overline{\Phi}(U) = \overline{\Psi}(U) = \frac{ND\left(1-\delta_2\right)^3}{3t|U|} + \frac{2ND\delta_2}{t|U|}.$$

**Доказательство** полностью аналогично доказательству теоремы 6.

**Следствие 2** *Пусть выполняются условия теоремы 11. Тогда*

$$\overline{\Phi}(U) = \underline{\Phi}(U) + \frac{2ND\delta_2}{t|U|},$$

$$\overline{\Psi}(U) = \underline{\Psi}(U) + \frac{2ND\delta_2}{t|U|}.$$

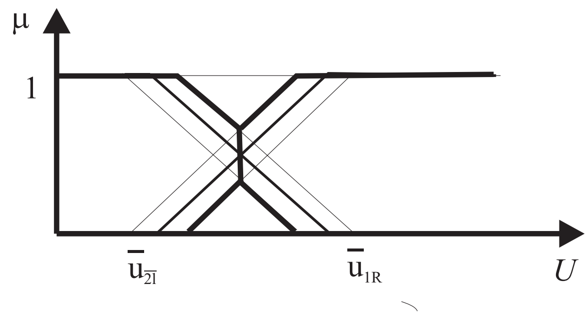


Рис. 3.7 Нижние оценки $\underline{\nu}(l_t)$ в $\delta-$ модели.

**Степень нечеткости в $\delta$ - модели**

Построим обобщение формулы ( 2.6) для случая $\delta-$ модели. Для этого рассмотрим некоторую точку $u \in U$ (рис. 3.3). Достаточно очевидно, что что нижние и верхние оценки для степени нечеткости в точке $\underline{\nu}(l_t)$ и $\overline{\nu}(l_t)$ будут достигаться при функциях принадлежности, имеющих изображенный на рис. 3.7 и 3.8 вид соответственно (на рис. 3.7 изображена наиболее близкая к характеристической функция принадлежности; на рис. 3.8 - наиболее близкая к функции $\mu(u) = 0.5 \forall u \in U$ функция принадлежности из данной области).

Для формализации записи таких функций введем следующие величины:

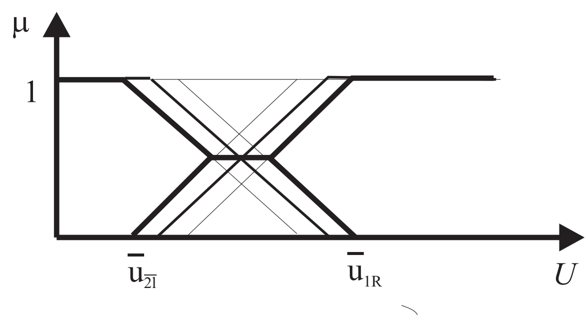


Рис. 3.8 Верхние оценки $\overline{\nu}(l_t)$ в $\delta-$ модели.

$$q = \{R, L\}, \tag{3.70}$$

$$\overline{q} = \begin{cases} R, & \text{если } q = L, \\ L, & \text{если } q = R, \end{cases} \tag{3.71}$$

$$\widetilde{U} = \cup_{j=2}^{t} \left[ \frac{\overline{u}_{jL} + \overline{u}_{j-1,R}}{2} - \frac{\delta_1}{2}, \frac{\overline{u}_{jL} + \overline{u}_{j-1,R}}{2} + \frac{\delta_1}{2} \right] \tag{3.72}$$

Тогда

$$\underline{\eta}(l_t, u) = 1 - \left( \mu_{i_1^*}^{q}(u) - \mu_{i_2^*}^{q}(u) \right), \tag{3.73}$$

где

$$\mu_{i_1^*}^{q}(u) = \max_{1 \le j \le t} \left\{ \mu_j^R(u), \mu_j^L(u) \right\}, \tag{3.74}$$

$$q = \begin{cases} R, & \mu_{i_1^*}^{R}(u) \ge \mu_{i_1^*}^{L}(u) \\ L, & \mu_{i_1^*}^{L}(u) > \mu_{i_1^*}^{R}(u) \end{cases}$$

$$\mu_{i_2^*}^{q}(u) = \max_{1 \le j \le t_{j \ne i_1^*}} \mu_j^q(u).$$

$$\overline{\eta}(l_t, u) = \begin{cases} 1 - \left( \mu_{i_1^*}^{\overline{q}}(u) - \mu_{i_2^*}^{\overline{q}}(u) \right), & u \in U \backslash \widetilde{U} \\ 0.5, & u \in \widetilde{U} \end{cases} \tag{3.75}$$

Аналогично ( 2.6) нижние и верхние оценки $\nu(l_t)$ имеют следующий вид:

$$\underline{\nu}(l_t) = \frac{1}{|U|} \int_U \underline{\eta}(l_t, u) du, \tag{3.76}$$

$$\overline{\nu}(l_t) = \frac{1}{|U|} \int_U \overline{\eta}(l_t, u) du. \tag{3.77}$$

**Теорема 13** *Пусть функции принадлежностей источника информации и пользователя различаются в каждой точке универсума не больше, чем на* $\delta_2$ *и* $l_2 \in G_2(\overline{L})$*. Тогда*

$$\underline{\nu}(l_2) = \frac{d\left(1-\delta_2\right)^2}{2|U|}, \overline{\nu}(l_2) = \frac{d\left(1+2\delta_2\right)}{2|U|}.$$

**Доказательство.** Как видно из рис. 3.7 и формул ( 3.73), ( 3.74), можно представить $\underline{\nu}(l_2)$, аналогично доказательству теоремы 5, следующим образом:

$$\begin{aligned}
\underline{\nu}(l_2) &= \frac{1}{|U|}\int_U \underline{\eta}(l_2,u)du = \frac{1}{|U|}\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{1R}-\frac{\delta_1}{2}} \underline{\eta}(l_2,u)du = \\
&= \frac{2}{|U|}\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{2L}+\frac{d}{2}} \left(1-\left(\frac{1}{d}\left[\left(\overline{u}_{1R}+\frac{\delta_1}{2}\right)-u\right]-\right.\right. \\
&\qquad\qquad \left.\left.-\frac{1}{d}\left[u-\left(\overline{u}_{2L}+\frac{\delta_1}{2}\right)\right]\right)\right)du = \\
&= \frac{2}{|U|}\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{2L}+\frac{d}{2}} du + \frac{4}{d|U|}\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{2L}+\frac{d}{2}} \left(u-\overline{u}_{2L}\right)du - \\
&\qquad\qquad -\frac{2}{d|U|}\int_{\overline{u}_{2L}+\frac{\delta_1}{2}}^{\overline{u}_{2L}+\frac{d}{2}} \left(d-\delta_1\right)du = \\
&= \frac{2}{d|U|}\left[\left(\frac{d}{2}-\frac{\delta_1}{2}\right)+2\int_{\frac{\delta_1}{2}}^{\frac{d}{2}} zdz - \left(d+\delta_1\right)\left(\frac{d}{2}-\frac{\delta_1}{2}\right)\right] = \\
&= \frac{2}{d|U|}\left[\left(\frac{d}{2}-\frac{\delta_1}{2}\right)\left(d-d-\delta_1\right)+2\frac{1}{2}\left(\frac{d^2}{4}-\frac{\delta_1^2}{4}\right)\right] = \\
&= \frac{2}{d|U|}\left[\left(\frac{d}{2}-\frac{\delta_1}{2}\right)\left(-\delta_1\right)+\left(\frac{d}{2}+\frac{\delta_1}{2}\right)\left(\frac{d}{2}-\frac{\delta_1}{2}\right)\right] = \\
&= \frac{2}{d|U|}\left[\left(\frac{d}{2}-\frac{\delta_1}{2}\right)\left(\frac{d}{2}-\frac{\delta_1}{2}\right)\right] = \frac{2}{d|U|}\frac{\left(d-\delta_1\right)^2}{4} = \\
&= \frac{d\left(1-\delta_2\right)^2}{2|U|} \qquad (3.78)
\end{aligned}$$

При выводе ( 3.78) использовалась замена переменных $z =$

$u - \underline{u}_{2L}$ и соотношение $\delta_1 = d\delta_2$ ( 3.43).

Аналогично вычислению $\underline{\nu}(l_2)$, представим $\overline{\nu}(l_2)$ следующим образом:

$$\begin{aligned}
\overline{\nu}(l_2) &= \frac{1}{|U|}\int_U \overline{\eta}(l_2,u)du = \frac{1}{|U|}\int_{\overline{u}_{2L}-\frac{\delta_1}{2}}^{\overline{u}_{1R}+\frac{\delta_1}{2}} \overline{\eta}(l_2,u)du = \\
&= \frac{2}{|U|}\int_{\overline{u}_{2L}-\frac{\delta_1}{2}}^{\frac{\overline{u}_{2L}+\overline{u}_{1R}}{2}} \overline{\eta}(l_2,u)du = \\
&= \frac{2}{|U|}\int_{\overline{u}_{2L}-\frac{\delta_1}{2}}^{\frac{\overline{u}_{2L}+\overline{u}_{1R}}{2}-\frac{\delta_1}{2}} \left(1 - \left(\frac{1}{d}\left[\left(\overline{u}_{1R} - \frac{\delta_1}{2}\right) - u\right] - \right.\right. \\
&\quad \left.\left. -\frac{1}{d}\left[u - \left(\overline{u}_{2L} - \frac{\delta_1}{2}\right)\right]\right)\right)du + \frac{2}{|U|}\int_{\frac{\overline{u}_{2L}+\overline{u}_{1R}}{2}-\frac{\delta_1}{2}}^{\frac{\overline{u}_{2L}+\overline{u}_{1R}}{2}} du = \\
&= \frac{2}{d|U|}\left[\int_{\overline{u}_{2L}-\frac{\delta_1}{2}}^{\overline{u}_{2L}+\frac{d}{2}+\frac{\delta_1}{2}} d\,du + \int_{\overline{u}_{2L}-\frac{\delta_1}{2}}^{\overline{u}_{2L}+\frac{d}{2}-\frac{\delta_1}{2}} 2\left(u - \overline{u}_{2L}\right)du - \right. \\
&\quad \left. - \int_{\overline{u}_{2L}-\frac{\delta_1}{2}}^{\overline{u}_{2L}+\frac{d}{2}-\frac{\delta_1}{2}} \left(d - \delta_1\right)du\right] + \frac{2}{|U|}\frac{\delta_1}{2} = \\
&= \frac{2}{d|U|}\left[d\frac{d}{2} + 2\int_{-\frac{\delta_1}{2}}^{\frac{d}{2}-\frac{\delta_1}{2}} z\,dz - \left(d-\delta_1\right)\frac{d}{2}\right] + \frac{\delta_1}{|U|} = \\
&= \frac{2}{d|U|}\left[\frac{d}{2}\left(d - d - \delta_1\right) + \frac{d}{2}\left(\frac{d}{2} - \delta_1\right)\right] + \frac{\delta_1}{|U|} = \\
&= \frac{d}{2|U|} + \frac{\delta_1}{|U|} = \frac{d}{2|U|} + \frac{d\delta_2}{|U|} = \frac{d}{2|U|}\left(1 + 2\delta_2\right) \qquad (3.79)
\end{aligned}$$

При выводе ( 3.79) использовалась замена переменных и соотношение $\delta_1 = d\delta_2$ ( 3.43).

Теорема доказана.

**Следствие 3** *Пусть выполняются условия теоремы 9. Тогда*

$$\underline{\Phi}(U) = \underline{\Psi}(U) = \frac{N\left(1-\delta_2\right)}{3}\underline{\nu}(l_2).$$

Доказательство непосредственно следует из теорем 9 и 13.

**Следствие 4** *Пусть выполняются условия теоремы 9. Тогда*

$$\overline{\Phi}(U) = \overline{\Psi}(U) = \frac{N\left(6\delta_2 + (1-\delta_2)^3\right)}{3\,(1+2\delta_2)}\overline{\nu}(l_2).$$

Доказательство непосредственно следует из теорем 10 и 13.

Терема 13 довольно легко обобщается на случай более двух значений признака.

**Теорема 14** *Пусть функции принадлежности источника информации и пользователя различаются в каждой точке универсума не более, чем на* $\delta_2$ *и* $l_t \in G_2(\overline{L})$. *Тогда*

$$\underline{\nu}(l_t) = \frac{D\,(1-\delta_2)^2}{2|U|},$$

$$\overline{\nu}(l_t) = \frac{D\,(1+2\delta_2)}{2|U|},$$

*где* $D = \sum_{j=1}^{t-1} d_{j,j+1}$, $d_{j,j+1} = \overline{u}_{jR} - \overline{u}_{j+1,L}$.

Доказательство аналогичо доказательству теоремы 13.

**Следствие 5** *Пусть выполняются условия теоремы 11. Тогда*

$$\overline{\Phi}(U) = \overline{\Psi}(U) = \frac{2N}{t\,(1+2\delta_2)}\left[\frac{(1-\delta_2)^3}{3} + 2\delta_2\right]\overline{\nu}(l_t).$$

Доказательство следует из теорем 12 и 14.

**Следствие 6** *Пусть выполняются условия теоремы 11. Тогда*

$$\underline{\Phi}(U) = \underline{\Psi}(U) = \frac{2N}{3t}\,(1-\delta_2)\,\underline{\nu}(l_t).$$

Доказательство следует из теорем 11 и 14.

## Глава 4.

# Поиск по нечетким запросам в четких базах данных

Рассмотрим частный случай ситуации 4 (таблица 1.2), а именно случай, когда мы формулируем нечеткие запросы к четкой базе данных (ситуация 3 в таблице 1.2).

Эта ситуация является довольно распространенной. Действительно, накопленные базы данных, содержащие конкретную информацию по различным объектам являются повседневным инструментом в работе многих категорий специалистов. При общении с базами данных необходимо формулировать запрос на "языке", понятном этой базе данных. Абстрагируясь от конкретных поисковых средств, мы можем утверждать, что терм запроса должен содержать как минимум название атрибута и ограничения на его значения (иначе запрос будет просто не понятен для поисковых средств базы данных). Указанное ограничение на "свободу пользователя" при формулировании запросов для многих конкретных ситуаций не является слишком жестким. Когда такой подход не доставляет неудобств пользователю? Можно выделить следующие характеристики информационной среды, когда это удобно:

- Пользователь ищет *конкретный* объект: он точно знает значения признаков интересующего его объекта. Если такого объекта нет, поиск прекращается.

- Пользователь ищет объекты, *похожие на конкретный* объект: он точно знает значения признаков интересующих его объектов. Если такие объекты не найдены, пользователь может "расширить" ограничения на значения интересующих его признаков. При этом база данных должна быть достаточно "маленькой", чтобы незначительное такое расширение не приводило к выдаче большого числа описаний объектов, которые придется обрабатывать "вручную".

Данные требования к информационной среде выполняются, например, в ситуациях, когда технолог хочет узнать наличие на складе конкретных деталей или веществ; когда в базе данных ищется владелец конкретного автомобиля или адрес проживания конкретного человека и т.п. Иными словами, лингвистическая модель пользователя эквивалентна лингвистическому обеспечению базы данных: все запросы пользователя однозначно выражаются на языке поисковых средств используемой базы данных.

Как быть в ситуациях, когда лингвистическая модель пользователя и лингвистическая модель базы данных различаются? Приведем несколько примеров (может быть "бытовых") такой ситуации.

**Пример 3** *Пусть у нас есть база данных по автомобилям, содержащая следующую информацию:*

*1. Марка;*

*2. Цена;*

*3. Год выпуска.*

*Мы хотим подобрать автомобиль для себя. В рамках традиционной технологии поиска мы должны указать интересующую нас марку (марки, связанные оператором "или"), цену (диапазон цен), год выпуска (диапазон годов выпуска). При этом нам выдаются те и только те описания, которые удовлетворяют заданным ограничениям. Вполне возможна ситуация, когда наиболее подходящий нам автомобиль по некоторому параметру лежит за пределами сформулированных ограничений (например, дороже на несколько тысяч рублей), и его описание, естественно, не выдается. Попытки расширить диапазон поиска при достаточно "плотной" базе данных могут привести к выдаче большого дополнительного числа описаний (информационный шум), которые придется обрабатывать "вручную".*

*Нам же необходим* дешевый, камфортабельный, не очень старый *автомобиль или* не очень дорогой, престижный, довольно новый, автомобиль для поездки по городу *и т.п. Как "объяснить" базе данных такого типа понятия пользователя? Как осуществлять по ним поиск информации?*

Описываемая ниже технология позволяет формализованно определять такого рода понятия пользователя и производить по ним поиск информации в "обычных" базах данных.

## 4.1. Описание нечеткого лингвистического интерфейса

Нечеткий лингвистический интерфейс (НЛИ) имеет 3 блока:

- блок формализации понятий пользователя;
- блок поиска информации;
- блок анализа результатов поиска.

## 4.2. Блок формализации понятий пользователя

*Блок формализации понятий пользователя* позволяет ввести новый поисковый признак и множество его значений. Например, к рассматриваемой в примере 3 базе данных можно ввести признак "Камфортабельность" со значениями "камфортабельный" "средней кафортабельности", "мало камфортабельный"; признак "Цена" со значениями "дорогой", "средней стоимости", "дешевый"; признак "Назначение" со значениями "для поездок по городу", "для поездок за город", "универсал" и т.п.

Далее, мы должны "объяснить" базе данных "смысл" введенных понятий - значений поискового признака. В соответствии с Заде [10], под смыслом некоторого понятия понимается его объем, то есть множество реальных объектов с указанием степени их принадлежности к понятию. Другими словами, задание смысла понятия эквивалентно заданию его функции принадлежности. При этом в качестве универсального множества выступает множество значений некоторого атрибута базы данных (домен). Продолжая рассмотрение примера 3, в качестве универсума признака "Камфортабельность" должно быть все множество марок автомобилей, которое есть в нашей базе данных;

в качестве универсума признака "Цена" - отрезок от минимальной цены в базе данных до максимальной цены.

Существует много различных методов задания функций принадлежности. Классификация таких методов и их краткий обзор приведены в приложении В. Вообще говоря, выбор того или иного метода зависит от задачи, существующей ситуации (например, наличия экспертов) и других параметров. В разработанных в ходе выполнения НИР макетах были использованы простейшие такие методы - непосредственное задание функций принадлежности.

Отметим, что при построении функций принадлежности мы можем, при выполнении некоторых условий, построить оптимальное множество значений признака с точки зрения пользователя - множество, в рамках которого неопределенность при формировании запроса будет минимальна 2.

Итак, выход блока формализации понятий пользователя - набор значений признака с указанием функций принадлежности, им соответствующих. Эти функции принадлежности определены на доменах соответствующих атрибутов базы данных.

## 4.3. Блок поиска информации

*Блок поиска информации по нечетким запросам* реализует следующий алгоритм поиска информации по нечетким понятиям пользователя.

1. Выбирается первая запись в базе данных.

2. Берется первый поисковый признак, сформулированный в запросе, и его значение.

3. По признаку выбирается соответствующий атрибут базы данных.

4. Вычисляется значение функции принадлежности, соответствующей значению признака, в точке, соответствующей значению анализируемого атрибута в анализируемой записи базы данных. Полученное значение функции принадлежности запоминается.

5. Берется следующий поисковый признак и повторяются шаги 3 и 4. Повторение происходит до тех пор, пока не кончаются поисковые признаки. В результате мы имеем набор значений

функций принадлежности значений атрибутов анализируемой записи нечеткому запросу.

6. На базе полученного в результате выполнения шага 5 набора значений вычисляется обобщенная оценка принадлежности анализируемой записи запросу. Процедура такого вычисления может опираться на изложенные в разделе 1 настоящего отчета результаты. В частности, это может минимальное значение из указанного набора, произведение элементов набора и т.п. Полученное обобщенное значение запоминается в рабочем поле базы данных.

7. Выбирается следующая запись в базе данных и повторяется шаг 2. Повторение происходит до тех пор, пока не переберутся все записи в базе данных.

Таким образом, результат поиска по нечеткому запросу - упорядочивание записей в базе данных по степени их соответствия данному запросу от 1 (полное соответствие) до 0 (полное несоответствие).

## 4.4. Блок анализа результатов поиска

*Блок анализа результатов поиска* предназначен для уточнения результатов поиска информации. Допуская неопределенность на входе НЛИ, мы должны предоставить средства ее "снятия" для конкретного пользователя на "выходе". Функции принадлежности, построенные в блоке формализации, отражают либо обобщенное мнение экспертов, либо мнение какого-либо эксперта в зависимости от выбранного метода построения функций принадлежности. Если мнение пользователя и мнение эксперта (экспертов) не совпадают, необходимо предоставить механизм "индивидуализации" функций принадлежности. Этот механизм устроен следующим образом.

Если пользователь не удовлетворен результатом поиска, он может сформулировать коррекцию запроса по конкретному признаку в виде направления модификации и модификатора. Направление модификации указывает полюс признака (например, "дешевле", "дороже", "престижнее", "менее престижное" и т.п.), модификатор - его "силу" (например, "чуть", "слегка", "значительно" и др.). По направлению модификации вычисляется направление сдвига функций принадлежности используемых понятий, по модификатору - шаг сдвига. Шаг сдвига может вычисляться разными способами, и этот вопрос требует отдельного исследования. В реализованных алгоритмах макета этот наг зависит от количества объектов, находящихся медджу существующей точкой и концом универсума. Если таких объектов много - шаг выбирается достаточно большим, если мало - достаточно маленьким.

Полученные в результате модификации функции принадлежности снова подаются на вход блока поиска информации. Таким образом, пользователь может "уточнять" результаты поиска информации.

Итак, НЛИ позволяет:

- в водить пользователю понятия, не содержащиеся в описаниях объектов базы данных;
- определять "смысл" введенных понятий как нечетких подмножеств доменов атрибутов базы данных;
- осуществлять поиск информации по введенным понятиям; результат поиска - упорядочивание записей базы данных по степени их соответствия запросу пользователя;
- уточнять результаты поиска информации за счет указания направления модификации и модификатора смысла используемых понятий.

В результате пользователь получает "лингвистическую оболочку" базы данных. Наличие различных классов пользователей, имеющих различные функции принадлежности, позволяет создавать и хранить несколько таких оболочек одной и той же базы данных, или несколько различных "взглядов" на одну и ту же информацию. Этот подход позволит значительно повысить эффективность использования существующих баз данных (см. введение к главе 2, раздел  4).

**Прил. А.**

# Доказательства теорем

## А.1. Доказательство леммы 1

Напомним определение расстояния.

**Определение 1** *Пусть $V$ - некоторое множество. $d(x, y)$ -* расстояние *в $V$, если $\forall x, y, z \in V$ :*
$1) d(x, y) \geq 0$
$2) d(x, x) = 0$
$3) d(x, y) = d(y, x)$
$4) d(x, z) \leq d(x, y) + d(y, z)$.

Для доказательства леммы необходимо проверить выполнение аксиом расстояния (определение 1).

Так как $\rho(\mu_j, \mu'_j)$ - метрика, то выполнение первых двух аксиом для ( 2.2) очевидно. Проверим выполнение аксиом 3, 4.

3. $d(s_t, s'_t) = 0 \Leftrightarrow \forall j (1 \leq j \leq t) \quad \rho(\mu_j, \mu'_j) = 0 \Leftrightarrow$
$\Leftrightarrow \forall j (1 \leq j \leq t) \quad \mu_j = \mu'_j \Leftrightarrow s_t = s'_t$.

4. $d(s_t, s'_t) = \sum\limits_{j=1}^{t} \rho(\mu_j, \mu'_j) \leq \sum\limits_{j=1}^{t} [\rho(\mu_j, \mu''_j) + \rho(\mu''_j, \mu'_j)] =$
$= \sum\limits_{j=1}^{t} \rho(\mu_j, \mu'_j) + \sum\limits_{j=1}^{t} \rho(\mu'_j, \mu''_j) = d(s_t, s''_t) + d(s^{''}_t, s^{'}_t)$.

## А.2. Доказательство теоремы 1

1. Выполнение аксиомы $A1$ очевидно.

2. Обозначим через $\eta(s_t, u)$ подинтегральную функцию в ( 2.4) В этом случае:

$$\begin{aligned} \xi(s_t) = 0 \quad & \Leftrightarrow \forall u \in U \; \eta(s_t, u) = 0 \Leftrightarrow \forall u \in U \; \mu_{i_1^*}(u) - \mu_{i_2^*}(u) = 1 \Leftrightarrow \\ & \Leftrightarrow \forall u \in U \exists j (1 \leq j \leq t) : \; \mu_j(u) = 1, \; \mu_i(u) = 0 \quad \forall i \neq j. \end{aligned}$$

3. Аналогично пункту 2,

$$\begin{aligned}\xi(s_t)=1 \quad & \Leftrightarrow \forall u\in U\ \eta(s_t,u)=1 \Leftrightarrow \forall u\in U\ \mu_{i_1^*}(u)-\mu_{i_2^*}(u)=0 \Leftrightarrow\\ & \Leftrightarrow \forall u\in U\ \exists i_1,i_2 (1\le i_1,i_2\le t):\\ & \quad \mu_{i_1}(u)=\mu_{i_2}(u)=\max\nolimits_{1\le j\le t}\mu_j(u).\end{aligned}$$

4. Рассмотрим произвольные $s_t,\ s'_{t'}\in G_t(L)$.

$$\begin{aligned}
\xi(s_t)\le\xi(s'_{t'}) \quad & \Leftrightarrow \quad \tfrac{1}{|U^1|}\textstyle\int_{U^1} f(\mu^1_{i_1^*}(u^1)-\mu^1_{i_2^*}(u^1))du^1 \le\\
& \le \quad \tfrac{1}{|U^2|}\textstyle\int_{U^2} f(\mu^2_{i_1^*}(u^2)-\mu^2_{i_2^*}(u^2))du^2 \Leftrightarrow\\
& \Leftrightarrow \quad \tfrac{1}{|U^1|}\textstyle\int_{U^1}(\mu^1_{i_1^*}(u^1)-\mu^1_{i_2^*}(u^1))du^1 \ge\\
& \ge \quad \tfrac{1}{|U^2|}\textstyle\int U^2(\mu^2_{i_1^*}(u^2)-\mu^2_{i_2^*}(u^2))du^2 \Leftrightarrow\\
& \Leftrightarrow \quad \tfrac{1}{|U^1|}\textstyle\int_{U^1}(1-(\mu^1_{i_1^*}(u^1)-\mu^1_{i_2^*}(u^1)))du^1 \le\\
& \le \quad \tfrac{1}{|U^2|}\textstyle\int_{U^2}(1-(\mu^2_{i_1^*}(u^2)-\mu^2_{i_2^*}(u^2)))du^2 \Leftrightarrow\\
& \Leftrightarrow \quad \tfrac{1}{|U^1|}\textstyle\int_{U^1}\Big[(1-\mu^1_{i_1^*}(u^1))+(\mu^1_{i_2^*}(u^1)-0)\Big]du^1 \le\\
& \le \quad \tfrac{1}{|U^2|}\textstyle\int_{U^2}\Big[(1-\mu^2_{i_1^*}(u^2))+(\mu^2_{i_2^*}(u^2)-0)\Big]du^2 \Leftrightarrow\\
& \Leftrightarrow \quad \tfrac{1}{|U^1|}\textstyle\int_{U^1}\Big[(h^1_{i_1^*}(u^1)-\mu^1_{i_1^*}(u^1))+\\
& \qquad\qquad +(\mu^1_{i_2^*}(u^1)-h^1_{i_2^*}(u^1))\Big]du^1 \le\\
& \le \quad \tfrac{1}{|U^2|}\textstyle\int_{U^2}\Big[(h^2_{i_1^*}(u^2)-\mu^2_{i_1^*}(u^2))+\\
& \qquad\qquad +(\mu^2_{i_2^*}(u^2)-h^2_{i_2^*}(u^2))\Big]du^2.
\end{aligned}$$

В данных преобразованиях вторая эквивалентность является следствием определения функции $f$ (требование $F2$), третья эквивалентность - следствие неравенства $0\le\mu_{i_1^*}(u)-\mu_{i_2^*}(u)\le$

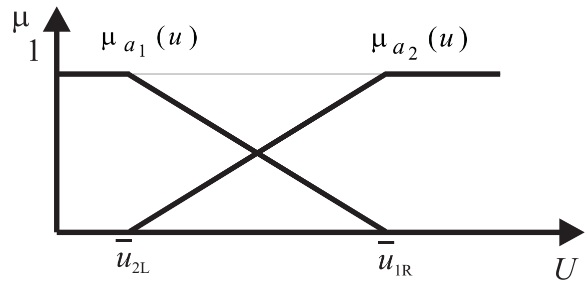


Рис. А.1 Графическое изображение $s_2 \in G_2(\overline{L})$.

1 $\forall u \in U$, замена 1 и 0 на $h$ с соответствующими индексами (последняя эквивалентность) - следствие определений $h$ ( 2.3) и $\mu_{i_1^*}(u),\ \mu_{i_2^*}(u)$ ( 2.5). Последнее неравенство мы можем переписать следующим образом.

$$d(h^1_{i_1^*},\ \mu^1_{i_1^*}) + d(h^1_{i_2^*},\ \mu^1_{i_2^*}) \leq d(h^2_{i_1^*},\ \mu^2_{i_1^*}) + d(h^2_{i_2^*},\ \mu^2_{i_2^*}),$$

где $d(h, \mu) = \int_U |h(u) - \mu(u)| du$ мера в $L$.

Последнее неравенство есть расстояние между $s_t$ и $\tilde{s}_t$ (Лемма 1).

Итак, при выполнении условий теоремы существует мера, для которой

$\xi(s_t) \leq \xi(s'_{t'})$, если $\rho(s_t, \tilde{s}_t) \leq \rho(s'_{t'}, \tilde{s'_{t'}})$.

## А.3. Доказательство теоремы 2

Рассмотрим простейший случай $t = 2$ (Рис. А.1).

Зафиксируем две точки:

$\overline{u}_{2L}$ - левая ненулевая граница $\mu_{a_2}(u)$ и

$\overline{u}_{1R}$ - правая ненулевая граница $\mu_{a_1}(u)$.

Значение интеграла ( 2.6) не равно нулю только на отрезке $[\overline{u}_{2L}, \overline{u}_{1R}]$. Таким образрм,

$$\xi(s_2) = \frac{1}{|U|}\int_U (1-(\mu_{i_1^*}(u)-\mu_{i_2^*}(u)))du = $$
$$= \frac{1}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}} (1-(\mu_{i_1^*}(u)-\mu_{i_2^*}(u)))du. \qquad (A.1)$$

Используя простейшие формулы элементарной геометрии, мы можем написать:

$$\mu_{a_1}(u) = \begin{cases} 1, & \text{если } u \le \overline{u}_{2L} \\ \frac{1}{d}(\overline{u}_{1R}-u), & \text{если } \overline{u}_{2L} \le u \le \overline{u}_{1R} \\ 0, & \text{если } u \ge \overline{u}_{1R} \end{cases}, \qquad (A.2)$$

$$\mu_{a_2}(u) = \begin{cases} 0, & \text{если } u \le \overline{u}_{2L} \\ \frac{1}{d}(u-\overline{u}_{2L}), & \text{если } \overline{u}_{2L} \le u \le \overline{u}_{1R} \\ 1, & \text{если } u \ge \overline{u}_{1R} \end{cases}, \qquad (A.3)$$

где $d = \overline{u}_{1R} - \overline{u}_{2L}$.

Подставляя ( A.2), ( A.3) в ( 2.6) и вспоминая ( 2.5), мы можем написать:

$$\xi(s_2) = \frac{1}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}} (1-(\mu_{i_1^*}(u)-\mu_{i_2^*}(u)))du =$$
$$= \frac{1}{|U|}\Big[\int_{\overline{u}_{2L}}^{\frac{1}{2}(\overline{u}_{1R}+\overline{u}_{2L})} (1-(\mu_{i_1^*}(u)-\mu_{i_2^*}(u)))du +$$
$$+ \int_{\frac{1}{2}(\overline{u}_{1R}+\overline{u}_{2L})}^{\overline{u}_{1R}} (1-(\mu_{i_1^*}(u)-\mu_{i_2^*}(u)))du\Big] =$$
$$= \frac{1}{2}|U|\int_{\overline{u}_{2L}}^{\frac{1}{2}(\overline{u}_{1R}+\overline{u}_{2L})} (\overline{u}_{1R}+\overline{u}_{2L}-2u)du =$$
$$= \frac{\overline{u}_{1R}+\overline{u}_{2L}}{2|U|} = \frac{d}{2|U|}.$$

Для доказательства теоремы в общем случае $t > 2$ мы должны повторить наши рассуждения для всех областей неопределенности $[\overline{u}_{j,L}, \overline{u}_{j-1,R}]$ $(2 \le j \le t)$.

## А.4. Доказательство теоремы 4

Рассмотрим простейший случай $t = 2$ (смотри доказательство теоремы 2). Если $s_2 \in G_2(L)$, то, в силу ортогональности,

$$\mu_{a_1}(u) = 1 - \mu_{a_2}(u) \quad \forall u \in U$$

Таким образом,

$$\begin{aligned}
\xi(s_2) &= \tfrac{1}{|U|}\int_{\overline{u}_{2L}}^{\overline{u}_{1R}}(1-(\mu_{i_1^*}(u)-\mu_{i_2^*}(u)))du = \\
&= \tfrac{1}{|U|}\Big[\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})}(1-(\mu_{a_1}(u)-\mu_{a_2}(u)))du + \\
&\quad + \int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}}(1-(\mu_{a_2}(u)-\mu_{a_1}(u)))du\Big] = \\
&= \tfrac{2}{|U|}\Big[\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})}(1-((1-\mu_{a_2}(u))-\mu_{a_2}(u)))du + \\
&\quad + \int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}}(1-((1-\mu_{a_1}(u))-\mu_{a_1}(u)))du\Big] = \\
&= frac2|U|\Big[\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})}\mu_{a_2}(u)du + \int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}}\mu_{a_1}(u)du\Big]
\end{aligned} \tag{A.4}$$

Повторяя данные рассуждения для $s_t' = g(s_t)$, мы можем написать:

$$\xi(s_2') = \frac{2}{|g(U)|}\left[\int_{g(\overline{u}_{2L})}^{g(\mu_{a_1}^{-1}(\frac{1}{2}))}\mu_{a_2}'(u')du' + \int_{g(\mu_{a_1}^{-1}(\frac{1}{2}))}^{g(\overline{u}_{1R})}\mu_{a_1}'(u')du'\right] \tag{A.5}$$

Производя замену переменных $u' = g(u)$, мы можем переписать ( А.5) следующим образом:

$$
\begin{aligned}
\xi(s_2') &= \tfrac{2}{|g(U)|}\Big[\textstyle\int_{g(\overline{u}_{2L})}^{g(\mu_{a_1}^{-1}(\frac{1}{2}))} \mu_{a_2}'(g(u))dg(u)+ \\
&\qquad + \textstyle\int_{g(\mu_{a_1}^{-1}(\frac{1}{2}))}^{g(\overline{u}_{1R})} \mu_{a_1}'(g(u))dg(u)\Big] = \\
&= \tfrac{2}{|g(U)|}\Big[\textstyle\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})} \mu_{a_2}(u)g'(u)du+ \\
&\qquad + \textstyle\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)g'(u)du\Big]
\end{aligned}
\tag{A.6}
$$

Последнее соотношение в ( A.6) является следствием определения $g(s_t)$.

Равенство $\xi(s_2) = \xi(g(s_2))$ эквивалентно равенству

$$
\begin{aligned}
&\tfrac{2}{|U|}\left[\textstyle\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})} \mu_{a_2}(u)du + \int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)du\right] = \\
&= \tfrac{2}{|g(U)|}\left[\textstyle\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})} \mu_{a_2}(u)g'(u)du + \int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)g'(u)du\right]
\end{aligned}
\tag{A.7}
$$

Равенство ( A.7) мы можем переписать как

$$
\begin{aligned}
&\textstyle\int_{\overline{u}_{2L}}^{\mu_{a_1}^{-1}(\frac{1}{2})} \mu_{a_2}(u)\left(\tfrac{1}{|U|} - \tfrac{g'(u)}{|g(U)|}\right)du+ \\
&\qquad + \textstyle\int_{\mu_{a_1}^{-1}(\frac{1}{2})}^{\overline{u}_{1R}} \mu_{a_1}(u)\left(\tfrac{1}{|U|} - \tfrac{g'(u)}{|g(U)|}\right)du = 0
\end{aligned}
\tag{A.8}
$$

Если $g(u)$ - линейная функция, то $g(u) = ku+a$, где $k$, $a$ некоторые константы. Значение $g'(u) = k = \frac{g(u_2)-g(u_1)}{u_2-u_1} \quad \forall u_2,\ u_1 \in U$, и, в частности, $g'(u) = \frac{|g(U)|}{|U|}$. Используя последнее равенство, мы можем написать:

$$
\frac{1}{|U|} - \frac{g'(u)}{|g(U)|} = 0.
$$

Таким образом, ( A.8) справедливо.

Для доказательства теоремы в общем случае $t > 2$ мы должны повторить наши рассуждения для всех областей неопределенности $[\overline{u}_{j,L}, \overline{u}_{j-1,R}] \quad (2 \le j \le t)$.

Теорема доказана.

Прил. В.

# Методы построения функций принадлежности

При разработке систем обработки нечеткой информации необходимо каким-либо образом задавать функции принадлежности используемых понятий. Вообще говоря, этот вопрос находится за пределами теории нечетких множеств и больше относится к теории экспертного оценивания и методов обработки экспертной информации. Однако, без краткого обзора таких методов представление о теории нечетких множеств, и, особенно ее прикладных аспектах, не будет достаточно полным. Поэтому приводимый в данной главе обзор имеет своей целью составить общее представление о методах построения функций принадлежности, и, при необходимости, указать литературу, которая может помочь при решении данной проблемы в конкретных ситуациях.

Существует ряд методов построения функций принадлежности нечеткого множества. Можно предложить следующую классификацию таких методов.

## В.1. Классификация методов построения функций принадлежности

По характеру измерений методы построения функций принадлежности можно разбить на прямые и косвенные.

Прямые методы определяются тем, что эксперт непосредственно задает правила определения значений функций принадлежности , характеризующее понятие $A$. Эти значения согласуются с его предпочтениями на множестве объектов $U$ следующим образом:

1) для любых $u_1, u_2 \in U$ $\mu_A(u_1) < \mu_A(u_2)$ тогда и только тогда, когда $u_2$ предпочтительнее $u_1$, т.е. в большей степени

характеризуется понятием $A$;

2) для любых $u_1, u_2 \in U$ $\mu_A(u_1) = \mu_A(u_2)$ тогда и только тогда, когда $u_1$ и $u_2$ базразличны относительно понятия $A$.

В качестве примеров прямых методов можно привести непосредственное задание функций принадлежности таблицей, формулой, примером [52], [59], [62]. К прямым методам можно также отнести методы, основанные на вероятностной трактовке функций принадлежности $\mu_A(u) = P(A|u)$, т.е. вероятности того, что объект $u \in U$ будет отнесен к множеству, которое характеризует понятие $A$.

В косвенных методах значения функций принадлежности выбираются таким образом, чтобы удовлетворять заранее сформулированным условиям. Информация от экспертов является только исходной информацией для дальнейшей обработки. Дополнительные условия могут налагаться как на вид получаемой информации, так и на процедуру дальнейшей ее обработки. Примерами дополнительных условий могут служить следующие:

- функция принадлежности должна отражать близость к заранее выделенному эталону;

- объекты множества $U$ являются точками в параметрическом пространстве [58];

- результатом процедуры обработки должна быть функция принадлежности, удовлетворяющая условиям интервальной шкалы [8];

- при попарном сравнении объектов, если один из них оценивается в $\alpha$ раз сильнее, чем другой, то второй объект оценивается в $frac1\alpha$ раз сильнее, чем первый [55];

- при определении степени принадлежности множество исследуемых объектов должно содержать, по крайней мере, два объекта, численные представления которых на интервале [0,1] равны 0 и 1 соответственно [59].

Выбор тех или иных методов диктуется условиями конкретной задачи. Общие рекомендации по их использованию могут быть сформулированы следующим образом.

Если гарантируется, что в данной задаче эксперты работают как надежные и правильные "измерители", то их можно спрашивать непосредственно о значениях приназдежности (прямые методы), если такого утверждать нельзя, то более подходящими являются косвенные методы. Косвенные методы более трудоемки, чем прямые, но более стойки по отношению к искажениям в ответах.

Свойство "измеримости" понятий также может быть существенным при выборе метода построения функции принадлежности. Как правило, прямые методы используются для описания понятий, которые характеризуются измеримыми свойствами, такими как высота, рост, объем и т.п. В этом случае удобно непосредственное задание значений степени принадлежности. Если мы рассмотрим более сложное понятие (например, "красота"), то окажется, что практически не существует универсальных элементарных измеримых свойств, описывающих данное понятие. В таких случаях используются только ранговые измерения при попарном сравнении объектов.

Функция принадлежности может отражать как мнение одного эксперта, так и мнение группы экспертов. Следовательно, возможны четыре группы методов: прямые и косвенные для одного эксперта и прямые и косвенные для группы экспертов.

## B.2. Прямые методы для одного эксперта

Прямые методы для одного эксперта, предлагающие непосредственное назначение степени принадлежности или назначение аналитической функции, совпадающей с функцией принадлежности, обсуждаются в [52], [59], [62], [32].

В [59] анализируется предложенный Осгудом [48] метод семантических дифференциалов для описания понятия посредством нечеткого множества характеризующих его свойств. Суть этого метода заключается в следующем:

1) определяется список свойств, по которым оценивается понятие (объект);

2) для каждого свойства формируется полярная шкала;

3) для каждой шкалы оценивается, как сильно описываемое понятие (объект) обладает положительным свойством;

4) формируется профиль понятия как совокупность оценок по шкалам.

**Пример 4** *В задаче распознавания лиц (фоторгафий) можно выделить следующие шкалы:*

| | *Название (имя) шкалы* | *Полюса шкалы* |
|---|---|---|
| $x_1$ | *высота лба* | *низкий (узкий) - широкий* |
| $x_2$ | *профиль носа* | *горбатый - курносый* |
| $x_3$ | *длина носа* | *короткий - длинный* |
| $x_4$ | *разрез глаз* | *узкие - широкие* |
| $x_5$ | *цвет глаз* | *темные - светлые* |
| $x_6$ | *форма подбородка* | *треугольный - квадратный* |
| $x_7$ | *толщина губ* | *тонкие - толстые* |
| $x_8$ | *цвет лица* | *смуглое - белое* |
| $x_9$ | *очертание лица* | *овальное - квадратное* |

*Таким образом, каждому лицу (фотографии) мы можем поставить в соответствие нечеткое подмножество универсального множества*

$$U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$$

В [52], [57] приводится следующая модель построения функций принадлежности. Пусть $\mathcal{K}$ - покрытие универсума $U$, т.е. совокупность обычных подмножеств $\{A_1, \cdots, A_s\}$ таких, что:

1)$A_i \in U \forall i (1 \leq i \leq s)$;

2)$A_i \neq \forall i (1 \leq i \leq s)$;

3)$A_1 \cap \cdots \cap A_s = U$.

Пусть $B \subset U$. Тогда $B$ может рассматриваться как нечеткое подмножество $\mathcal{K}$ с функцией принадлежности

$$\mu_B(A_i) = \frac{A_i \cap B}{A_i \cup B},$$

где $|A_i|$ - число элементов (мощность) множества $A_i$.

**Пример 5** *Пусть* $U = \{1, \cdots, 9\}, A_1 = \{1, 3, 5\}, A_2 = \{3, 6, 9\}, A_3 = \{2, 4, 8\}, A_4 = \{1, 3, 7\}, A_5 = \{2, 3, 8\}, B = \{2, 3, 5, 8, 9\}$. *Тогда* $\mu_B(A_i) = \{1/3, 1/3, 1/3, 1/7, 3/5\}$.

## B.3. Косвенные методы для одного эксперта

Как отмечалось выше (раздел B.1), в практике чато имеют место случаи, когда не существует элементарных измеримых свойств, через которые определяются интересующие нас понятия (например, понятия "интеллектуальность", "красота" и т.п.). В таких случаях трудно проранжировать степень проявления свойства у анализируемых объектов (например, в процедуре Осгуда - раздел B.2). В подобных случаях используют процедуру парного срвнения объектов.

Идея данного метода, предложенного Саати в [56], заключается в следующем. Если бы значения степени принадлежности некоторого понятия $S$ на универсуме $U = \{u_1, \cdots, u_n\}$ были известны ($\mu_S(u_i) = w_i, i = 1, 2, \cdots, n$), то результат парных сравнений можно было бы представить матрицей отношений $A = (a_{ij})$, где $a_{ij} = frac w_i w_j, i, j = 1, 2, \cdots, n$. Если данные отношения известны точно, то получается соотношение $Aw = nw$, где $w = (w_1, \cdots, w_n), n$ - собственное значение матрицы $A$, по которому можно восстановить вектор $w$. Так как отношения сравнения $a_{ij}$ в реальном случае не точны (как и любые эмпирические данные), то мы должны вычислить оценку для $w$. Эта оценка вычисляется при следующих предположениях:

1)$a_{ii} = 1 (1 \leq i \leq n)$, т.е. любой объект максимально похож сам на себя;

2)$a_{ij} = \frac{1}{a_{ji}} (1 \leq i, j \leq n)$, т.е. если один из объектов оценивается в $\alpha$ раз сильнее, чем другой, то второй объект оценивается в $\frac{1}{\alpha}$ раз сильнее, чем первый.

В этом случае

$$\sum_{j=1}^{n} \frac{a_{ij} w_j}{w_i} = n (i = 1, 2, \cdots, n),$$

где $n$ - наибольшее собственное значение $A$, а другие собственные значения $\lambda$ равны 0, так как

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii} = n.$$

В общем случае эмпирическая шкала $w = (w_1, \cdots, w_n)$ должна удовлетворять задаче на поиск собственного значения $Aw = \lambda_{\max}$, где $\lambda_{\max}$ - наибольшее собственное значение $A$. Следовательно, задача сводится к поиску вектора $w$, который удовлетворяет уравнению $Aw = \lambda_{\max} w$. Чем более ближе $\lambda_{\max}$ к числу $n$, тем более верным является результат. Так как известно, что задача $Aw = \lambda_{\max} w$ имеет единственное решение, то значения координат наибольшего собственного вектора, деленные на их сумму, будут искомыми степенями принадлежности.

В [55] описывается применение данного метода для анализа сложных свойств, которые представляются как иерархическая система.

В [34] предлагается метод наименьших квадратов для получения значений функции принадлежности по матрице бинарных отношений $A = (a_{ij})$. Искомые значения получаются как решение оптимизационной задачи:

$$f = \sum \sum (a_{ij} w_j - w_i)^2 \to \min, \sum w_i = 1, w_i > o.$$

Иллюстрируется близость оценок относительных степеней принадлежности, полученных данными методами.

В [3] предлагается процедура построения функций принадлежности, в которой используется матрица парных сравнений с неполной информацией (то есть допускается, что некоторые ее элементы отсутствуют). Рассматривается понятие $''S''$, которое является функцией принадлежности на множестве объектов $A = \{a_1, \cdots, a_n\}$. Предполагается, что в $A$ имеется два объекта $a^0$ и $a^1$, таких, что $a^1$ - идеальный представитель тех объектов, которые принадлежат $''S''$ (то есть $\mu_S(a^1) = 1$) и $a^0$ - идеальный представитель тех объектов, которые не принадлежат $''S''$ (то есть $\mu_S(a^0) = 0$). Эксперту предлагается проранжировать степень различия объектов в каждой паре объектов по принадлежности к понятию $''S''$. В результате формируется матрица парных сравнений, которая задает порядок пар объектов по степени различия в парах. Далее посредством методов неметрического шкалирования вычисляются в факторном (метрическом) пространстве $X^m$ координаты $n$ точек $x^i = \{x^i_1, \cdots, x^i_m\} (i = 1, 2, \cdots, n)$, порядок расстояний $d(x^i, x^j)$

между которыми совпадает или максимально близок к порядку элементов матрицы парных сравнений. Доказывается, что для полученных расстояний имеют место следующие утверждения:

1) если объекты $a_i$ и $a_j$ неразличимы, то $d_{ij} = d(a_i, a_j) = 0$;

2) если степень различия объектов $a_i$ и $a_j$ больше, чем степнь различия объектов $a_i$ и $a_k$, то $d_{ij} > d_{ik}$;

3) если степень различия объектов $a_i$ и $a_j$ совпадает со степнью различия объектов $a_i$ и $a_k$, то $d_{ij} = d_{ik}$.

Далее предполагается, что степень различия двух объектов $a_i$ и $a_j$ по отношению к понятию $''S''$ будет пропорциональна разности значений функций принадлежности на этих объектах, то есть

$$|d_{1i} - d_{1j}| = |\mu_S(a_i) - \mu_S(a_j)|,$$

где - некоторая константа.

Рассматривая в качестве объекта $a_i$ объекты $a^0$ и $a^1$ соответственно, получаем:

$$c(d_{10} - d_{1j}) = \mu_S(a_j), cd_{1j} = 1 - \mu_S(a_j).$$

Из данных уравнений непосредственно следует, что

$$\mu_S(a_j) = \frac{d_{10} - d_{1j}}{d_{1j}} = 1 - \frac{d_{1j}}{d_{10}}.$$

Таким образом, функция принадлежности определяется по расстояниям в пространстве признаков $X^m$.

В [58] описывается общий метод варьирования прототипов для получения численного значения функций принадлежности. Этот подход используется в [33] для распознавания образов.

# B.4. Прямые методы для группы экспертов

Прямые методы для группы экспертов объединяют все методы, разработанные в рамках обработки экспертной информации (например, [15]) применительно к конкретному вопросу - определению значений функций принадлежности. Достаточно полный обзор всех таких методов требует введения многих специфических и не нужных далее понятий и поэтому занимает слишком много места. В то же время, выбор конкретного метода зависит от решаемой задачи, поэтому целесообразно ограничиться присанием одного из них, иллюстрирующем основные идеи этой группы методов.

Одной из интерпретаций понятия "функция принадлежности" является так называемая вероятностная интерпретация [52]. В [4] описана следующая методика оценки значений функции принадлежности, базирующаяся на такой интерпретации.

При построении функции принадлежности некоторого понятия $A$ на универсуме $U$ рассматривается не только само понятие, но и его дополнение в $U$, обозначаемое $\neg A$. Сумма значений функций принадлежностей произвольного элемента $u \in U$ к системе таких понятий будет равна единице. Далее экспертам предлагается оценить в процентах (или в других единицах) степень проявления каждого понятия ($A$ и $\neg A$) в данной точке $u \in U$. Разработанные в рамках теории экспертных опросов методы позволяют оценить качество таких оценок (оценка согласованности, точности и т.п.).

Однако, в некоторых случаях мнение эксперта очень трудно выразить в процентах, поэтому более приемлимым способом оценки функции принадлежности может быть метод опроса, предложенный в [22]. Оцениваемое значение $u \in U$ предъявляется большому числу экспертов. Каждый эксперт имеет один голос. Он должен однозначно ответить на вопрос о принадлежности $u$ к классу $A$ и $\neg A$, т.е. возможны только ответы "Да, $u$ принадлежит $A$", "Да, $u$ принадлежит $\neg A$" (ответы тапа "Не знаю", "Затрудняюсь ответить" не допускаются). Значение функции принадлежности вычисляется по формуле

$$\mu_A(u) = \frac{n_A}{n},$$

где $n$ - число экспертов, учавствовавших в эксперименте, $n_A$ - число экспертов, ответивших: "Да, $u$ принадлежит $A$".

## В.5. Косвенные методы для группы экспертов

В [28] предлагается метод определения функции принадлежности на основе интервальных оценок. Пусть у нас есть $m$ экспертов ($m > 1$) и выделены $n$ признаков оцениваемого понятия $S$. Пусть интервал $[x_{ji}, x'_{ji}]$ отражает мнение $i-$ го эксперта о значении $j-$ го признака ($i = 1, \cdots, m; j = 1, \cdots, n$). Тогда полным описанием анализируемого понятия $i-$ экспертом является гиперпараллелепипед

$$\Theta_i = [x_{1i}, x'_{1i}] \times \cdots \times [x_{ni}, x'_{ni}].$$

Эта информация обрабатывается следующим образом.

1) Рассматривая для каждого признака $j$ все интервалы, предложенные экспертами, находим связное покрытие их объединения, состоящее из непересекающихся интервалов, концами которых являются только концы исходных интервалов:

$$[x_{jk}, x'_{jk}](j = 1, \cdots, n; k = 1, \cdots, m_j - 1).$$

2) Образуем на основе полученных покрытий непересекающиеся гиперпараллелепипеды

$$T_k = [x_{1k}, x'_{1k}] \times \cdots \times [x_{nk}, x'_{nk}].$$

3) Вычисляем для $x \in T_k$ функции

$$\phi_i(x) = \begin{cases} 0, & \text{при} \quad T_k \cap \Theta_i \neq \emptyset; \\ 1, & \text{при} \quad T_k \cap \Theta_i = \emptyset. \end{cases}$$

4) Полагаем номер итерации $l = 1$.

5) Вводим коэффициенты компетентности $\lambda_i^l = \frac{1}{m}(i = 1, \cdots, m)$.

6) Вычисляем $l$ - приближение функции принадлежности по формуле

$$f^l(x) = \sum_{i=1}^{m} \phi_i(x)\lambda_i^l, x \in T_k.$$

7) Вычисляем функционал рассогласования мнения $i-$ эксперта с мнением экспертного совета на $l-$ итерации по формуле

$$\delta_i^l = \sum_{x \in T_k} \left[ f^l(x) - \phi_i(x) \right]^2, i = 1, \cdots, m.$$

8) Вычисляем $\Delta = \sum_{i=1}^m \frac{1}{\delta_i^l}$.

9) Присваиваем $l = l + 1$.

10) Вычисляем $\lambda_i^l = \frac{\Delta}{\delta_i^{l-1}}$.

11) Если $\max |\lambda_i^{l-1} - \lambda_i^l| < \epsilon$, то считаем $\mu_S(x) = f^l(x)$. Иначе возвращаемся к шагу 6).

В [12] предлагается следующий метод для группы экспертов. Каждый эксперт $E_i (i = 1, \cdots, m)$ выделяет из $U$ множество элементов $Q_i$, соответствующих, по его мнению, понятию $S$. Далее рассматривается подмножество $U \ni Q = \bigcup_{i=1}^m Q_i$. Каждым экспертом ранжируются все элементы $Q$ по степени соответствия понятию $S$. Эксперт может использовать отношения эквивалентности ($\approx$) или порядка ($\succ, \succeq$). Предполагается также, что эксперты могут поставить коэффициетны степени предпочтения $\gamma$ перед элементами в упорядоченной последовательности, усиливая или ослабляя отношение предпочтения. Для каждого такого упорядочивания, данного $i-$ экспертом, вводится расстояние между его элементами с порядковыми номерами $r$ и $s$ следующим образом.

1) Вычисляется расстояние между соседними элементами цепочки по формуле

$$\rho(q_j, q_{j+1}) = \begin{cases} 1, & \text{при} \quad q_j \succ q_{j+1} \\ \frac{1}{2}, & \text{при} \quad q_j \succeq q_{j+1}, \\ 0, & \text{при} \quad q_j \approx q_{j+1} \end{cases} \quad (1 \leq j \leq |Q|);$$

2) Вычисляется расстояние между первым элементом упорядочивания и $s-$ элементом по формуле

$$\rho_s^i = \rho(q_1^i - q_s^i) = \sum_{j=r}^{s-1} \gamma_j \rho(q_j^i, q_{j+1}^i) (2 \leq j \leq |Q|);$$

3) Вычисляется расстояние между $r$ и $s$ элементами по формуле

$$\rho(q_r^i, q_s^i) = \rho(q_1^i, q_s^i) - \rho(q_1^i, q_r^i) = \rho_s^i - \rho_r^i.$$

Последняя разность показывает, насколько предпочтительнее $q_r$ по сравнению с $q_s$ с точки зрения $E_i$.

Далее предполагается, что разность между весами $\phi(q_r^i) - \phi(q_s^i)$ пропорциональна разности $\rho_s$ и $\rho_r$. При $r = s+1$ получаем рекуррентную формулу

$$\phi(q_{s+1}^i) - \phi(q_s^i) = c(\rho_{s+1}^i - \rho_s^i).$$

Таким образом, задача сводится к определению веса первого элемента $\phi(q_1^i)$. На основе всех $\phi(q_r^i)(i = 1, \cdots, m)$ для $q_r$ определяется значение

$$\phi(q_r) = \frac{1}{m}\sum_{i=1}^{m}\phi(q_r^i).$$

Последнее и есть степень принадлежности элемента $u \in U$ нечеткому множеству $S$.

## В.6. Параметрическое задание функций принадлежности

Описываемые ниже способы задания функций принадлежности не являются, собственно, методами их построения. Это некоторые способы представления функций принадлежности, отражающие интуицию и опыт ученых и инженеров, использующих данный аппарат.

В [9] предлагается следующее параметрическое задание функций принадлежности. Вводятся два типа функций: $s-$ функции и $\pi-$ функции, где

$$S(u;\alpha,\beta,\gamma) = \begin{cases} 0 & \text{для} \quad u \le \alpha, \\ 2\left(\frac{u-\alpha}{\gamma-\alpha}\right)^2 & \text{для} \quad \alpha \le u \le \beta, \\ 1 - 2\left(\frac{u-\gamma}{\gamma-\alpha}\right)^2 & \text{для} \quad \beta \le \ u \le \gamma, \\ 1 & \text{для} \quad u \ge \gamma. \end{cases} \qquad (В.1)$$

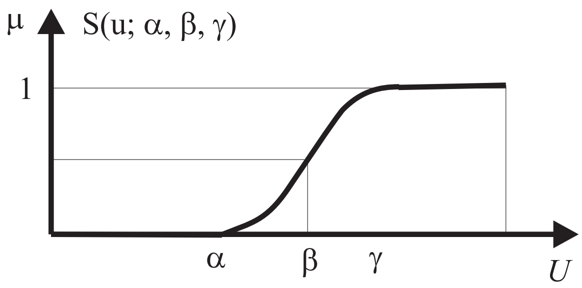


Рис. В.1 Графическое изображение $s-$ функции.

$$\pi(u;\beta,\gamma)=\begin{cases} S(u;\gamma-\beta,\gamma-\frac{\beta}{2},\gamma) & \text{для} \quad u\le\gamma, \\ S(u;\gamma,\gamma+\frac{\beta}{2},\gamma+\beta) & \text{для} \quad u\ge\gamma. \end{cases} \qquad (\text{B.2})$$

(Рис. В.1 и В.2 соотвественно).

Параметры данных формул могут "вычисляться" одним из описанных выше методов. Такое представление функций принадлежности оказалось удобным для решения ряда задач (например, арифметические операции над нечеткими числами

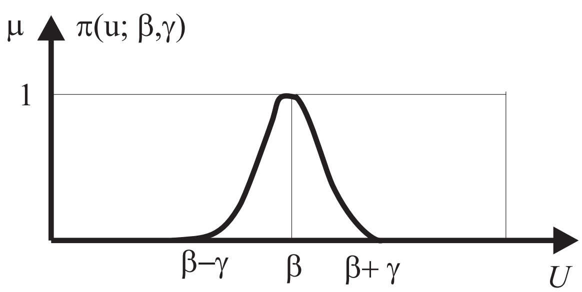


Рис. В.2 Графическое изображение $\pi-$ функции.

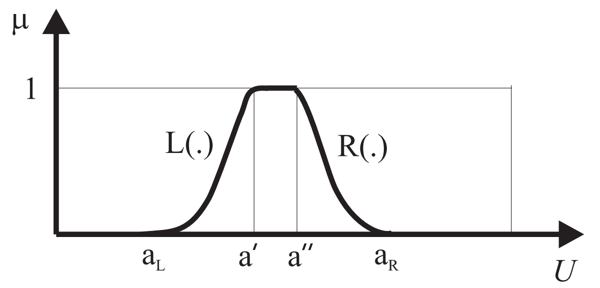


Рис. В.3 Графическое изображение функции принадлежности $(L-R)-$ типа.

можно выразить как операции над данными параметрами), однако, не получило большого использования в приложениях. Последнее связано с неоправданной грмоздкостью данного представления, трудностями хранения и обработки подобных функций в компьютерах и специализированных аппаратных средствах.

Позже было предложено аппроксимировать подобные "плавные" параметричесие функции кусочно - линейными функциями (функциями принадлежности $(L-R)-$ типа - рис. В.3).

Это оказалось чрезвычайно удобным для хранения и обработки и послужило основой для создания развития специализированных аппаратно-программных средств обработки нечеткой информации (например, [54], [30], [53], [39]. Заметим, что соответствующие параметры могут определяться одним из описанных выше методов построения функций принадлежности.

В заключение приведем примеры параметрического задания конкретных функций принадлежности [14].

**Пример 6** *Функции принадлежности утверждения* "величина $u$ мала".

*1)* $\mu(u) = \begin{cases} 1 & \text{\textit{при}} \quad 0 \le u \le a; \\ 0 & \text{\textit{при}} \quad u > a. \end{cases}$

*2)* $\mu(u) = e^{-ku}, k > 0.$

*3)* $\mu(u) = e^{-ku^2}, k > 0.$

*4)* $\mu(u) = \begin{cases} 1 - au^k & \text{при} \quad 0 \le u \le \frac{1}{\sqrt[k]{a}}; \\ 0 & \text{при} \quad \frac{1}{\sqrt[k]{a}} \le u. \end{cases}$

*5)* $\mu(u) = \frac{1}{1+ku^2}, k > 1.$

*6)* $\mu(u) = \begin{cases} 1 & \text{при} \quad 0 \le u \le a; \\ \frac{1}{2} - \frac{1}{2}\sin\frac{\pi}{b-a}\left(u - \frac{a+b}{2}\right) & \text{при} \quad a \le u \le b; \\ 0 & \text{при} \quad b \le u. \end{cases}$

*6)* $\mu(u) = \begin{cases} 1 & \text{при} \quad 0 \le u \le a; \\ \frac{b-u}{b-a} & \text{при} \quad a \le u \le b; \\ 0 & \text{при} \quad b \le u. \end{cases}$

**Пример 7** *Функции принадлежности утверждения* "величина $u$ большая".

*1)* $\mu(u) = \begin{cases} 0 & \text{при} \quad 0 \le u \le a; \\ 1 & \text{при} \quad u > a. \end{cases}$

*2)* $\mu(u) = \begin{cases} 0 & \text{при} \quad 0 \le u \le a; \\ 1 - e^{-k(u-a)} & \text{при} \quad a \le u, \end{cases}, k > 0.$

*3)* $\mu(u) = \begin{cases} 0 & \text{при} \quad 0 \le u \le a; \\ 1 - e^{-k(u-a)^2} & \text{при} \quad a \le u, \end{cases}, k > 0.$

*4)* $\mu(u) = \begin{cases} 0 & \text{при} \quad 0 \le u \le a; \\ b(u-a)^k & \text{при} \quad a \le u \le c, c = a + \frac{1}{\sqrt[k]{b}}; \\ 1 & \text{при} \quad c \le u. \end{cases}$

*5)* $\mu(u) = \begin{cases} 0 & \text{при} \quad 0 \le u \le a; \\ \frac{k(u-a)^2}{1+k(u-a)^2} & \text{при} \quad a \le u. \end{cases}$

*6)* $\mu(u) = \begin{cases} 0 & \text{при} \quad 0 \le u \le a; \\ \frac{1}{2} + \frac{1}{2}\sin\frac{\pi}{b-a}\left(u - \frac{a+b}{2}\right) & \text{при} \quad a \le u \le b; \\ 1 & \text{при} \quad b \le u. \end{cases}$

*6)* $\mu(u) = \begin{cases} 0 & \text{при} \quad 0 \le u \le a; \\ \frac{u-a}{b-a} & \text{при} \quad a \le u \le b; \\ 1 & \text{при} \quad b \le u. \end{cases}$

# Литература


1. Артемьев В. Что такое Business Intelligence? *Открытые системы*, №4, 2003, с. 20 - 26.
2. Берг А.И. *Кибернетика - наука об оптимальном управлении.* М., "Энергия", 1964.
3. Блишун А.Ф. Моделирование процесса принятия решений в нечетких условиях на основе сходства понятий классов. Автореферат диссертации на соискание ученой степени кандидата физико-математических наук. М., ВЦ АН СССР, 1982, 19 с.
4. Борисов А.Н., Осис Я.Я. Методика оценки функций принадлежности элементов размытого множества. *Кибернетика и диагностика,* Рига, РПИ, 1970, с. 125-134.
5. Брутян Г.А. *Гипотеза Сепира - Уорфа.* Ереван, 1968, 120 с.
6. Галахов И. Проектирование корпоративной информационно - аналитической системы. *Открытые системы*, №4, 2003, с. 27 - 32.
7. Дюбуа , Прайд *1990 Теория возможностей: Приложения к представлению знаний в информатике.*
8. Жуковин В.Е., Оганесян Н.А., Бурштейн Ф.В., Корелов Э.С. Об одном подходе к задачам принятия решений с позиций теории нечетких множеств. *Методы принятия решений в условиях неопределенности.* Рига: РПИ, 1980, с. 12-16.
9. Заде Л.А. Размытые множества и их применение в распознавании образов и кластер-анализе. *Классификация и кластер.* Под. ред. Дж. Вэн Райзин, М., Мир, 1980, с. 208 - 247.
10. Заде Л.А. *Понятие лингвистической переменной и его применение к принятию приблизительных решений.* М., Мир, 1976, 165 с.
11. Ильин В.А., Поздняк Э.Г. *Основы математического анализа. Часть 1.* М. Наука, 1982, 616 с.
12. Киквидзе З.А., Ткемаладзе Н.Т. Об одном способе взвешивания элементов нечеткого множества. *Сообщения АН ГССР,* 1979, т. 93, № 2, с. 317-320.
13. В. В. Корнеев, А. Ф. Гареев, С. В. Васютин, В. В. Райх.

*Базы данных. Интеллектуальная обработка информации.* Москва, Нолидж, 2001, 352 с.

14. Кофман А. *Введение в теорию нечетких множеств. Пер. с франц.* М., Радио и связь, 1982, 432 с.
15. Литвак Б.Г. *Экспертная информация: методы получения и анализа.* М., Радио и связь, 1982. - 184 с.
16. Львов М. Построение информационно - аналитической системы. *Открытые системы*, №4, 2003, с. 39 - 42.
17. Ляпунов А.А. О некоторых общих вопросах кибернетики. *"Проблемы кибернетики",* вып.1, 1959.
18. *Нечеткие множества в моделях управления и искусственного интеллекта.* Под ред. Поспелова Д.А. М., Наука, 1986, 311 с.
19. *Построение экспертных систем.* Под редакцией Ф. Хейес-Рот, Д. Уотерман, Д. Ленат. М., Мир, 1987, 441 с.
20. Пфанцгаль И. *Теория измерений.* Пер. с англ. М., "Мир", 1978
21. Рыжов А.П. О степени нечеткости размытых характеристик. *Математическая кибернетика и ее приложения в биологии.* Под ред. Л.В.Крушинского, С.В.Яблонского, О.Б.Лупанова, М., Издательство МГУ, 1987, С. 60 - 77.
22. Рыжов А.П. Степень нечеткости лингвистической шкалы и ее свойства. *Нечеткие системы поддержки принятия решений.* Калинин, КГУ, 1989, с. 82-92.
23. Рыжов А.П. Степень нечеткости лингвистической шкалы и ее свойства. *Нечеткие системы поддержки принятия решений.* Под ред Аверкина А.Н. и др., Калинин, Издательство Калининского госуниверситета, 1988, С. 82 - 92 .
24. Стулов А. Особенности построения информационных хранилищ. *Открытые системы*, №4, 2003, с. 76 - 79.
25. Тэрано Т., Асаи К., Сугэно М. *1993 Прикладные нечеткие системы.*
26. Фу К. *Структурные методы в распознавании объектов.* Пер. с англ. М., "Мир", 1977.
27. Черняк Л. Что Business Intelligence предлагает бизнесу. *Открытые системы*, №4, 2003, с. 33 - 38.
28. Шер А.П. Согласование нечетких экспертных оценок и

функция принадлежности в методе размытых множеств. *Моделирование и исследование систем автоматического управления,* Владивосток, ДВНЦ АН СССР, 1978, с. 111-118.

29. Яблонский С.В. Основные понятия кибернетики. *"Проблемы кибернетики",* вып.2, 1959.
30. Bauer P., Klement E., Leikermoser A., Moser B. Approximation of Real Functions by Rule Bases. *Proceedings of the Fifth International Fuzzy Systems Association World Congress, Seoul, Korea, 1993,* V. 1, p. 239-242.
31. Bosc P., Piver O. *1993 On the evaluation of fuzzy quantified queries in a databases managment system.*
32. Braae M., Rutherford D.A. Selection of parametrs for fuzzy logic controller. *Fuzzy Sets and Systems,* 1979, V.2, p. 185-199.
33. Bremermann H. Pattern recognition. *Systems theory in the social sciences* Ed. by H. Bossel at al. Stuttgart: Binkhauser Verlag, 1976, p. 116-159.
34. Chu A.T.W., Kalaba R.E., Spingarn J. A comparison of two methods for determining the weights of belonging to fuzzy sets. *Journal of Optimization Theory and Applications,* 1979, V. 27, p. 531-538.
35. Dubois D., Prade H. Algorithmes de puls courts chemins pour traiter des données flous. *RAIRO. Recherche Operationelle*, 1978, v. 12, N 2, p. 213 - 227.
36. Dubois D., Prade H. Decision-making under fuzziness. *Advances in fuzzy set theory and applications*, Ed. by M.M. Gupta, R.K. Ragade, R.R.Jager, Amsterdam, North-Holland Publication Corporation, 1979, p. 279 - 302.
37. Fukami S., Mizumoto M., Tanaka K. Some considerations on fuzzy conditional inferences. *Fuzzy Sets and Systems,* 1980, v. 4, p. 243 - 273.
38. Kandel, Zemankova. Fuzzy relational data bases.
39. R. Krause, M. Schroder. An application of equality relations to idle speed control. - In: Proceedings of the First European Congress on Intelligent Technologies. September 7-10, 1993. Aachen, Germany. - V.1. - pp.370-376.
40. Mamdani E.H. Application of fuzzy ligic to approximate

reasoning using linguistic systems. *IEEE Transaktion Computational,* 1977, v. 26, p. 1182 - 1191.

41. Mansfield W.H., Fleischman R.M. *1993 A high perfomance fuzzy query processing system for relational databases.*
42. Michinori, Nakata *1991 Integrity constraints in fuzzy databases*
43. Mizumoto M., Fukami S., Tanaka K. Fuzzy conditional inferences and fuzzy inferences with fuzzy quantifiers. *Proceedings of the International Conference on Artificial Intelligence,* Tokyo, 1979, p. 20 - 23.
44. Mizumoto M., Tanaka K. Algebraic properties of fuzzy numbers. *Proceegings of IEEE International Conference on Cybernetisc and Society*, 1976, p. 559 - 563.
45. Mizumoto M., Tanaka K. Some properties of fuzzy numbers. *Advances in fuzzy set theory and applications*, Amsterdam, North-Holland Publication Corporation, 1979, p. 153 - 164.
46. Mizumoto M., Zimmermann H.J. Comparison of fuzzy reasoning methods. *Fuzzy Sets and Systems,* 1982, v.8, p. 253 - 283.
47. North American Fuzzy Logic Proceeding Society (NAFIPS'92). Proceedings of a Conference held in PUERTO VALLARTA, MEXICO, December 15-17, 1992. NASA Conference Publication 10112.
48. Osgood C.E., Suci G.L., Tannenbaum P.H. The measurement of meaning. *University of Illinois Press,* Urbana, 1957, p. 1-342.
49. Prade H. Using fuzzy set theory in scheduling problem: a case study. *Fuzzy Sets and Systems*, 1979, v. 2, N 2, p. 153 - 165.
50. Proceedings of the First European Congress on Intelligent Technologies. September 7-10, 1995. Aachen, Germany.
51. Proceedings of the Fifth International Fuzzy Systems Association World Congress'93. - July 4-9, 1993, Seoul, Korea.
52. Ragade R.K., Gupta M.M. Fuzzy sets theory: introduction. *Fuzzy Automata and Decision Processes* Ed. by Gupta M.M., Saridis G., Gaines B. Amsterdam: North-Holland, 1977, p. 105-131.
53. Raymond C., Boverie S., Le Quellec J.M. Practical realisation of fuzzy controllers comparison with conventional methods. *Proceedings of the First European Congress on Intelligent*

*Technologies. September 7-10, 1993. Aachen, Germany.* V.1. - pp.149-156.

54. Rueda A., Pedrycz W. Fuzzy Coordinator in Control Problems. *North American Fuzzy Logic Processing Society (NAFIPS'92). Proceedings of a Conference held in PUERTO VALLARTA, MEXICO, December 15-17, 1992,* NASA Conference Publication 10112, V. 1, p. 322-329.
55. Saaty T.L. Exploring the interface between hierarchies, multiple objectives and fuzzy sets. *Fuzzy Sets and Systems,* 1978, V.1, p. 57-69.
56. Saaty T.L. Measuring the fuzziness of sets. *Journal of Cybernetics,* 1974, V.4, p. 53-61.
57. Sanchez E. Inverses of fuzzy relations. Applications to possibility distribution and medical diagnosis. *Fuzzy Sets and Systems,* 1979, V.2, p.75-86.
58. Skala H.J. On many-valued logic, fuzzy logic and their applications. *Fuzzy Sets and Systems,* 1978, V.1, p. 129-149.
59. Thole U., Zimmermann H.J., Zysno P. On the suitability of minimum and products operators for the intersection of fuzzy sets. *Fuzzy Sets and Systems,* 1979, p. 1-14.
60. Zadeh L.A. Approximate reasoning in in fuzzy logic. *Proceedings of the International Conference on Artificial Intelligence,* Tokyo, 1979.
61. Zadeh L.A. A theory of approximate reasoning. *Machine Intelligence,* 1979, v.9, p. 149 - 194.
62. Zadeh L.A. Calculus of fuzzy restrictions. *Fuzzy Sets and Their Applications to Cognitive and Decision Processes,* Ed. by Zadeh L.A. and al. New York: Academic Press, 1975, p. 1-41.
63. Zadeh L.A. Fuzzy logic and approximate reasoning. *Synthese,* 1975, v. 80, p. 407 - 428.
64. Zadeh L.A. Fuzzy sets. *Information and Control*, 1965, v.8, p. 338 - 353.

# Предметный указатель


Ближайшая совокупность характеристических функций 19

Информационные шумы 29
  Верхняя оценка информационных шумов в $\delta$ - модели 50
  Нижняя оценка информационных шумов в $\delta$ - модели 50

Лингвистическая модель 69

Лингвистическая переменная 12
  Базовые термины 13
  Модификаторы 13

Лингвистические базы данных 26

Методика выбора оптимального множества значений качественного признака 44

Нечеткий запрос 10

Нечеткая база данных 10

Нечеткая переменная 12

Нечеткий лингвистический интерфейс 70
  Блок анализа результатов поиска 72
  Блок поиска информации по нечетким запросам 71
  Формализация понятий пользователя 70

Полное ортогональное семантическое пространство 18

Потери информации 29
  Верхняя оценка потерь информации в $\delta$ - модели 50
  Нижняя оценка потерь информации в $\delta$ - модели 50

Семантическое пространство 15
  Нормальность 17
  Полнота 17
  Ортогональность 18

Среда поиска информации 9

Степень нечеткости ПОСП 19
  Аксиомы степени нечеткости ПОСП 19

Интерпретация степени нечеткости ПОСП 21
Теорема существования 20
Верхняя оценка степени нечеткости в $\delta$ - модели 61
Нижняя оценка степени нечеткости в $\delta$ - модели 61

Четкая база данных 9

Четкий запрос 9

$\delta$ - модель 46

Для заметок

Для заметок

Рыжов Александр Павлович

Модели поиска информации в нечеткой среде.
М., Издательство Центра прикладных исследований при механико-математическом факультете МГУ, 96 стр.

*Оригинал макет изготовлен издательской группой механико-математического факультета МГУ*

Подписано в печать 10.03.2004 г.
Формат 60×90 1/16. Объем 6,0 п.л.
Заказ 2 Тираж 300 экз.

---

Издательство ЦПИ при механико–математическом факультете МГУ
г. Москва, Воробьевы горы.
Лицензия на издательскую деятельность ИД № 04059 от 20.02.2001 г.

---

Отпечатано на типографском оборудовании механико-математического факультета и Франко-русского центра им. А. М. Ляпунова