Методы получения, представления и обработки знаний с НЕ-факторами

В монографии рассматриваются три последовательных процесса работы со знаниями — получение, представление и обработка, причём акцент сделан на так называемых НЕ-факторах, то есть факторах неопределённости, которые обычно присутствуют в знаниях экспертов. Приводится обзор современных методов, подходов и технологий извлечения, представления и обработки таких знаний, даётся богатый список специализированной литературы.

Работа будет интересной студентам и аспирантам, обучающимся по специальности «искусственный интеллект», а также всем, кто живо интересуется этой темой.

ДУШКИН Роман Викторович

УДК 004.82 + 004.832.34
ББК 32.81
Д86

Душкин Р. В.
Д86 Методы получения, представления и обработки знаний с НЕ-факторами. — 2011. — 115 с., ил.

В монографии рассматриваются три последовательных процесса работы со знаниями — получение, представление и обработка, причём акцент сделан на так называемых НЕ-факторах, то есть факторах неопределённости, которые обычно присутствуют в знаниях экспертов. Приводится обзор современных методов, подходов и технологий извлечения, представления и обработки таких знаний, даётся богатый список специализированной литературы.

Работа будет интересной студентам и аспирантам, обучающимся по специальности «искусственный интеллект», а также всем, кто живо интересуется этой темой.

УДК 004.82 + 004.832.34
ББК 32.81

Методы получения, представления и обработки знаний с НЕ-факторами

ДУШКИН Роман Викторович
roman.dushkin@gmail.com

Москва, 2011

Принимаются благодарности

Вниманию всех читателей! Данная книга издана в электронном виде и распространяется абсолютно бесплатно. Вы можете свободно использовать её для чтения, копировать её для друзей, размещать в библиотеках на сайтах в сети Интернет, рассылать по электронной почте и при помощи иных средств передачи информации. Вы можете использовать текст книги частично или полностью в своих работах при условии размещения ссылок на оригинал и должном цитировании.

При этом автор будет несказанно рад получить читательскую благодарность, которая позволит как улучшить текст данной книги, так и более качественно подойти к подготовке следующих книг. Благодарности принимаются на счёт в платёжной системе «Яндекс.Деньги», на который также можно перечислить малую лепту и при помощи терминалов:

4100137733052

Убедительная просьба; по возможности, при перечислении благодарности указывать в пояснении к переводу наименование книги или какое-либо иное указание на то, за что именно выражается благодарность.

Содержание

Предисловие

В бытность аспирантом на кафедре Кибернетики в Московском Государственном Инженерно-физическом Институте (МИФИ) а также во времена моей работы там же мною было собрано, систематизировано, обобщено некоторое количество источников и материалов, касающихся манипуляции знаниями с так называемыми НЕ-факторами. Здесь под не совсем корректным термином «манипуляция» подразумеваются все те процессы, которые вынесены в заголовок настоящей монографии — получение, представление и обработка знаний. На основе собранной информации и имеющихся методов были разработаны и проверены практикой формализмы представления и процедуры обработки знаний с НЕ-факторами, причём НЕ-факторы могли проявляться одновременно. Полученные результаты должны были лечь в основу кандидатской диссертации, однако не сложилось.

Тем не менее, разработанные методики и полученные результаты, в частности, лингвистический подход к извлечению НЕ-факторов (основанный на квантификаторах), максиминный нечёткий вывод, некоторые методы дефаззификации и методы преобразования отдельных видов НЕ-факторов друг в друга, выглядят довольно интересными, чтобы быть представленными широкой публике, интересующейся методами искусственного интеллекта. В связи с чем было принято решение достать все собранные и разработанные материалы «из-под спуда» и сформировать на их основе небольшую монографию справочно-описательного характера.

И вот монография в руках у читателя. В ней читатель найдёт краткое введение и три главы. Во введении приведён необходимый минимум информации об используемых в книге понятиях, терминах, формализмах и методах. И три главы посвящены последовательно тем процессам, которые, опять же, вынесены в название монографии — в главах описываются методы получения знаний с НЕ-факторами, способы их представление, а также методы обработки таких знаний. От читателя ожидается определённый уровень подготовки — необходимо общее представление о формализмах дискретной математики, в первую очередь о теории множеств, теории вероятности и математической статистики.

Предисловие было бы неполным, если не описать в нём кратко о результатах использования описанных в настоящей монографии принципов и методов. Некоторое время назад под моим руководством выполнялся проект для Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий (МЧС России) в рамках выполнения Федеральной целевой программы «Повышение безопасности дорожного движения в 2006 — 2012 годах». Проект заключался в разработке Автоматизированной системы поддержки принятия решений (АСППР), в состав которой входила экспертная система, а сама АСППР была предназначена для использования в условиях ликвидации последствий дорожно-транспортных происшествий с участием транспортных средств, перевозящих опасные грузы (радиоактивно опасные, химически опасные и биологически опасные).

Упомянутая экспертная система из состава АСППР получала на вход данные от руководителя команды ликвидации последствий о визуально наблюдаемых факторах чрезвычайно ситуации, а также геоинформационные и метеорологические данные из смежной системы. На основе полученных входных данных рассчитывались рекомендации по используемым методикам, силам и средствам, которые необходимо задействовать для ликвидации последствий дорожно-транспортного происшествия.

Ситуация осложнялась тем, что зачастую по дорогам возят опасные грузы без какой-либо сопроводительной документации и с отсутствующими пометками на транспортных средствах (по крайней мере, так утверждали постановщики задачи). Поэтому острым стоял вопрос определения неизвестного вещества по визуально наблюдаемым факторам. Именно эта задача и стала той, на которой проверялись методы, описанные в настоящей монографии.

Собственно, задача является классической для использования аппарата нечёткой логики и вычислений в условиях неопределённости — на вход поступают нечёткие данные о визуально наблюдаемых свойствах опасных веществ, которые «утяжелены» факторами уверенности. Дополнительные ограничения, которые накладывались на решение задачи, заключались в том, что использование разрабатываемой экспертной системы должно было производиться в условиях, когда оператор одет в костюм полной химической защиты, так что приходилось решать ещё и побочную задачу грамотного отображения результатов машинного вывода и минимизации количества взаимодействия с носимым терминалом, на котором планировалась запускаться экспертная система.

В итоге была разработана и сдана МЧС России экспертная система, в которой была спроектирована и реализована методика определения неизвестного вещества, основанная на нечёткой логике и преобразовании неопределённости в нечёткость и наоборот. Данная система запрашивала у оператора минимальный набор значений визуально наблюдаемых свойств опасного вещества (цвет, агрегатное состояния, консистенция, плотность и несколько других), причём можно было ограничиться вводом только тех значений, которые в действительности наблюдались и поддавались оценке оператором. На основе введённой информации экспертная система давала рекомендации по типу опасного вещества, ранжированные по степени уверенности в них. Данные рекомендации в дальнейшем могли использоваться в методиках при расчёте сил и средств, необходимых для ликвидации последствий.

АСППР с описанной экспертной системой в её составе была успешно разработана и внедрена в Департаменте оперативного управления МЧС России в опытную эксплуатацию.

Надеюсь, что данный труд будет полезен тем специалистам в области искусственного интеллекта, а также студентам и аспирантам, занимающимся исследованиями в этом интереснейшем направлении.

Душкин Р. В.,
Москва, 2011.

Отзывы

«*Приятно сознавать, что у нас в стране ещё имеются молодые энтузиасты науки, которые своими собственными силами, без каких-либо грантов или формальных поощрений, пытаются продвигать научную и исследовательскую мысль. Предлагаемая книга будет полезна для тех, кто занимается искусственным интеллектом как на профессиональном, так и на любительском уровне. Без сомнения, книга станет ценным источником информации и знаний для студентов и аспирантов, обучающихся по специальностям «Прикладная математика» и «Информатика»*» —

Тарасов В. Б.,
член Научного совета Российской ассоциации искусственного интеллекта, вице-президент Российской ассоциации нечётких систем и мягких вычислений, доцент кафедры «Компьютерные системы автоматизации производства» МГТУ им. Н. Э. Баумана, к. т. н., доцент.

«*Сегодня катастрофически мало выходит доступной научной литературы в таких фундаментальных, но в то же время имеющих важное прикладное значение областях знаний, как искусственный интеллект. Собранный в монографии материал, равно как и описание собственных наработок автора, несомненно, помогут современным студентам и аспирантам, выбравшим для себя непростую стезю математических наук. Особую благодарность заслуживает труд автора за то, что он решил предоставить к данной книге бесплатный доступ*» —

Сергиевский Г. М.,
к. т. н., доцент кафедры «Системного анализа» НИЯУ МИФИ

«*Как мы знаем из истории, открытия, призванные изменить мир, редко даются безболезненно. И чем более многообещающе новое направление, чем значимее ожидаемый эффект, тем сложней и тернистей путь его исследователей. Так было и с исследованиями в области искусственного интеллекта. Бурный начальный период, когда первые успехи обещали прорыв в этой области, привлёк к теме искусственного интеллекта большое внимание общественности. Полёт фантазии уже рисовал фантастические картины нового мира, мира, где человечество получит себе в помощь Искусственный Разум. Поэтому первые трудности и поражения, показавшие нереальность этих ожиданий, были восприняты очень болезненно. В широких кругах термин «искусственный интеллект» стал почти ругательным, оттолкнув большое количество исследователей. И тем зна-*

чимее труд автора, не только продолжившего исследования этой темы, но и имеющего смелость вынести некоторые результаты своей работы на широкое обсуждение» —

Ершов А. В.,
технический директор Quintura Inc.,
автор 8-ми патентов США
по нейронным сетям.

«*Автор этой монографии известен большим опытом создания различных прикладных экспертных систем и их компонентов; с некоторыми из них, работающими в МИФИ, мне довелось работать лично. Он является специалистом по кибернетике и много сделал для популяризации этой науки: от обзоров книг различной степени сложности, до сборников энциклопедических статей, его авторства. Центральным понятием настоящей книги являются знания — именно в кибернетическом смысле — и теория «НЕ-факторов» это достаточно оригинальный подход, развитый в нашей стране несколькими крупными исследователями искусственного интеллекта. Она призвана решать трудности, возникающие при работе со знаниями в интеллектуальных системах, то есть развивать механизмы решения слабоформализованных задач. Вообще же, понятие «знания» сейчас всё больше нагружается научными смыслами как в искусственном интеллекте, так и в смежных областях. В настоящем труде, кстати, есть много указаний на достижения смежных областей (например, психологии), которые успел «переварить» управленческий подход к искусственному интеллекту. Несмотря на то, что сами психологи относятся к экспертным системам весьма скептически, это самый развитый и надёжный на сегодняшний день способ построения символьных систем для решения слабоформализованных задач*» —

Каунов С. А.,
Аспирант ВЦ РАН, член Российской
Ассоциации Искусственного Интеллекта.

Введение

Целью настоящей монографии является обзор современных подходов, методов и технологий получения, представления и обработки знаний с такими НЕ-факторами, как нечёткость, неопределённость, неточность и недоопределённость (определения терминов даны ниже в Глоссарии). Под получением понимаются механизмы (возможно разработанные только на теоретическом уровне) перевода информации, полученной от эксперта, из структурированных текстов или из баз данных, в некоторые формализованные структуры. Под представлением понимается форматирование формализованных структур, полученных на этапе извлечения, с целью долгосрочного хранения (например, в базах знаний). Под обработкой понимается не только машинный вывод на знаниях, полученных на этапе извлечения, но также и верификация этих знаний, то есть уже формализованных структур.

В области инженерии знаний к настоящему моменту разработано достаточно большое число методов, стратегий и процедур работы с экспертами, предложены различные способы обработки полученных в результате взаимодействия с экспертами результатов, а также создан целый ряд программных средств, автоматизирующих процессы извлечения знаний из экспертов, специальных текстов на естественном или структурированном языке и баз данных.

Однако, как показывает всесторонний анализ отечественных и зарубежных работ по проблемам приобретения знаний, практически малоисследованными остаются вопросы приобретения знаний при формировании непротиворечивых баз знаний с так называемыми НЕ-факторами [21, 22]. Эти вопросы имеют большое значение, поскольку знания, извлечённые из экспертов, как правило, содержат различные виды НЕ-факторов, в связи с чем соответствующие методы и процедуры приобретения знаний должны обеспечивать возможность извлечения и обработки не полностью известной (недостоверной) информации. В целом решением проблем, связанных с представлением и обработкой каждого из НЕ-факторов, занимаются самостоятельные направления исследований, где для этих целей создаются специальные математические аппараты и формализмы, наиболее известными из которых являются аппарат нечётких множеств [57, 23], методы недоопределённых моделей и так называемое программирование в ограничениях (*constraint programming*), а также теории возможностей и неопределённости (*belief and uncertainty theories*), на основе которых существует целый ряд конкретных подходов, в частности, описанные в работах [11, 37, 45, 55] и других. Тем не менее, вопросы применения теоретических результатов исследований в этой области к процессам автоматизированного приобретения знаний обсуждаются мало.

Соответственно структура монографии определяется рассмотренными факторами: имеется три главы, в каждой из которых описываются извлечение, представление и обработка соответственно. В каждой главе рассматриваются найденные в различных источниках формализмы, теории и технологии для четырёх выделенных НЕ-факторов: нечёткости, неопределённости, неточности и недоопределённости. В отдельных случаях предлагаются авторские подходы и методы к указанным процессам.

В первой главе описываются современные подходы к автоматизированному приобретению знаний с НЕ-факторами из экспертов и других источников знаний. Рассматриваются методы извлечения знаний с нечёткостью, неопределённостью, неточностью и недоопределённостью, границы их применимости, адекватность поставленным задачам.

Вторая глава содержит описания формализмов для представления выделенных НЕ-факторов. Рассматриваются примеры конкретного применения структур для представления знаний с НЕ-факторами.

В третьей главе представлены методы и технологии обработки НЕ-факторов — стратегии вывода на нечётких, неопределённых, неточных и недоопределённых знаниях, а также методы верификации знаний с этими НЕ-факторами. Приводятся примеры конкретных областей применений тех или иных методов, рассматривается сравнение некоторых методов с точки зрения их адекватности и простоты применения.

Глоссарий

Ниже приведены определения всех основных понятий и терминов, используемых в данной монографии. Где это возможно, приводятся различные трактовки и определения терминов, взятые из различных источников. Термины расположены не в алфавитном порядке, а в порядке значимости.

Знания

Знания — это самое кардинальное понятие технологии систем, основанных на знаниях [28]. За годы изучения проблемы специалистами было предложено множество различных толкований этого понятия через ряд специфических признаков, позволяющих соотнести его с понятием «данные». Сравнение обоих понятий приведено в следующей таблице:

Таблица 1. Сравнение структур знаний и данных

Знания (Зн)	Данные (Д)
$Зн_1$ — знания в памяти человека.	$Д_1$ — результат наблюдений над объектами или данными в памяти человека.
$Зн_2$ — материализованные знания (учебники, справочники).	$Д_2$ — фиксация данных на материальном носителе (таблицы, графики и т. д.).
$Зн_3$ — поле знаний (структурированное, полуформализованное описание $Зн_1$ и $Зн_2$).	$Д_3$ — модель данных (некоторая схема описания, связывающая несколько объектов).
$Зн_4$ — знания на языках представления знаний (формализация $Зн_3$).	$Д_4$ — данные на языке описания данных.
$Зн_5$ — база знаний в ЭВМ (на машинных носителях информации).	$Д_5$ — база данных на машинных носителях информации.
Традиционно выделяют три уровня: $Зн_1$ (знания) $\rightarrow$ $Зн_2$ (поле знаний) $\rightarrow$ $Зн_5$ (база знаний).	Традиционно выделяют три уровня: $Д_1$ (внешний) $\rightarrow$ $Д_3$ (логический) $\rightarrow$ $Д_5$ (физический).

Далее приводится совокупность качественных свойств для знаний, то есть специфических признаков знаний, позволяющих определить и охарактеризовать сам термин «знания»:

1. Знания имеют более сложную структуру, чем данные.
2. Знания задаются как экстенсионально (то есть через набор конкретных фактов, соответствующих рассматриваемому понятию), так и интенсионально (то есть через свойства, соответствующие рассматриваемому понятию), а данные всегда задаются экстенсионально.
3. Внутренняя интерпретируемость знаний — наличие возможности хранения в памяти совместно с элементом данных «избыточной» системы имён.
4. Рекурсивная структурированность знаний — наличие возможности расчленяться и объединяться по принципу «матрёшки».
5. Взаимосвязь (связанность) единиц знаний — наличие возможности установления различных отношений, отражающих семантику и прагматику связей отдельных явлений и фактов, а также отношений, отражающих смысл системы в целом.
6. Наличие у знаний семантического пространства с метрикой — возможность определять близость/удалённость информационных единиц.
7. Активность знаний — наличие возможности формировать мотивы поведения, ставить цели, строить процедуры их решения.
8. Функциональная целостность знаний — возможность выбора желаемого результата, времени и средств получения результата, средств анализа достаточности полученного результата.

На следующем рисунке показана обобщённая классификация знаний по [25]:

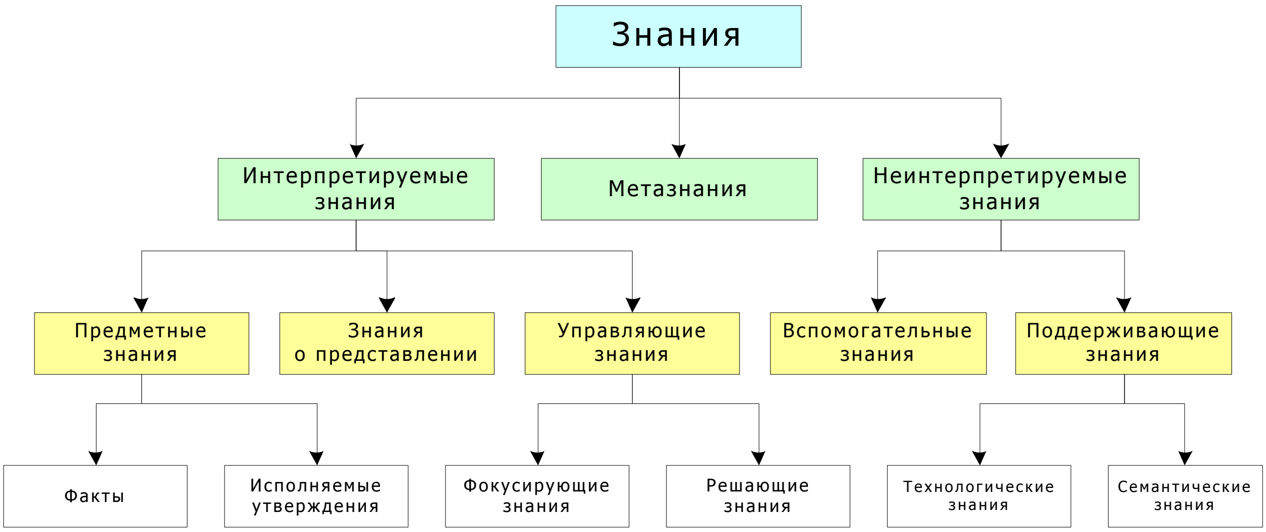

Рисунок 1. Обобщённая классификация знаний

Приобретение знаний

Под приобретением знаний понимается процесс получения знаний от эксперта или каких-либо других источников и формализация этих знаний для последующего использования их в системах, основанных на знаниях [28]. Приобретение знаний является одним из видов процесса получения знаний, при этом собственно под приобретением знаний понимается получение знаний из источников знаний при помощи использования программных средств поддержки деятельности инженера по знаниям и эксперта.

Другими двумя видами получения знаний является извлечение знаний (получение знаний из экспертов или других источников знаний без использования компьютерных средств поддержки этого процесса, а путём непосредственного контакта инженера по знаниям и источника знаний) и формирование знаний (получение знаний из источников при помощи использования программ обучения при наличии репрезентативной выборки примеров принятия решений в рассматриваемой предметной области).

Представление знаний

В соответствии с работой [25] представление знаний — это процесс, реализующий ответы на два вопроса: «*Что представлять?*» и «*Как представлять?*». Первый вопрос — это вопрос определения состава знаний, его важность определяется тем, что решение именно этой задачи обеспечивает адекватное отображение моделируемой проблемной области. Второй вопрос, в свою очередь, разделяется на две в значительной степени независимые задачи: как организовывать (структурировать) знания и как представить знания в выбранном формализме.

Необходимо отметить, что два главных вопроса представления знаний не являются независимыми друг от друга. Действительно, выбранный формализм представления может оказаться непригодным в принципе, либо неэффективным для выражения знаний о некоторых проблемных областях.

Обработка знаний

Здесь под обработкой знаний будет пониматься собственно вывод на продукционных правилах (см. «Машина вывода»), а также логическая проверка знаний, то есть верификация, хотя зачастую под последним процессом понимается также и синтаксическая и семантическая проверка полей и баз знаний.

Машина вывода (Решатель, Интерпретатор)

Один из трёх компонентов продукционной системы (экспертной системы, основанной на правилах вида «условие $\to$ действие»). Интерпретатор формально может быть представлен в виде четвёрки [25]:

$$I = \langle V, S, K, W \rangle,$$

где:

V — процесс выбора из базы знаний и из рабочей памяти продукционной системы подмножества активных продукций и подмножества активных данных, которые будут использованы на очередном цикле работы интерпретатора.

S — процесс сопоставления, определяющий множество означиваний, то есть множество пар: (правило p_i — данные d_j), при этом каждое правило p_i принадлежит подмножеству активных правил, а данные d_i являются подмножеством подмножества активных данных, полученных в процессе V.

K — процесс разрешения конфликтов (иначе — процесс планирования), определяющий, какое из означиваний будет выполняться.

W — процесс, осуществляющий выполнение выбранного правила для означивания в процессе K. Результатом выполнения является модификация рабочей памяти продукционной системы, либо операция ввода/вывода.

Верификация

В общем случае: процесс логической проверки полей и баз знаний. В случае использования продукционных систем процесс верификации знаний сводится к проверке и, возможно, устранению или исправлению определённых правил или наборов правил продукционной системы, выбранных по критериям верификации. К таким критериям относятся наличие определённых НЕ-факторов в правилах (противоречивые и неполные правила), логические ошибки в правилах (транзитивно-замкнутые правила), излишество правил и т. д. Кроме того, часто под верификацией понимается синтаксическая и семантическая проверка полей и баз знаний.

Продукция

В современном понимании термин продукция — это способ представления знаний в следующем наиболее общем виде [14]:

$$(i): Q; P; C; A \Rightarrow B; N,$$

где:

i — Собственное имя (метка) продукции.

Q — Сфера применения продукции, вычленяющая из предметной области некоторую её часть, в которой знания, заключённые в продукцию, имеют смысл.

P — Предусловие, содержащее информацию об истинности данной продукции, её приоритетности и т. п., используемое в стратегиях управления выводом для выбора данной продукции для исполнения.

C — Условие, представляющее собой предикат, истинное значение которого разрешает применять на некотором шаге данную продукцию.

$A \Rightarrow B$ — Ядро продукции. Интерпретация ядра продукции может быть различной, например: «если A истинно, то B истинно», «Если A — текущая ситуация, то надо делать B» и т. д.

N — Постусловие продукции, содержащее информацию о том, какие изменения надо внести в данную продукцию или другие продукции, входящие в систему продукций, после выполнения текущей продукции.

К достоинствам продукционного представления знаний относятся следующие:

1. *Модульность* — любая продукция может быть размещена в любом месте продукционной системы, так как организация знаний в продукционной системе обладает естественной модульностью. Поскольку каждая продукция — это законченный фрагмент знаний о предметной области, то всё множество продукций может быть разбито на подмножества, соответствующие описанию некоторого объекта.

2. *Единообразие структуры* (основные компоненты продукционной системы могут применяться для построения экспертных систем с различной проблемной ориентацией).

3. *Декларативность*, присущая продукционным системам, позволяет описывать предметную область, а не только строить программы преобразования информации. Кроме того, управление выводом и сам вывод осуществляется с использованием встроенного механизма.

4. *Естественность* (вывод заключения в продукционной системе во многом аналогичен процессу рассуждения эксперта).

5. *Независимость продукций* делает продукционные системы весьма перспективными для реализации на параллельных архитектурах, в частности, для разработки специализированных вычислительных комплексов, ориентированных на продукционные правила.

6. *Гибкость родовидовой иерархии понятий*, которая поддерживается только как связи между правилами (изменение правил влечёт за собой изменение и в иерархии).

7. *Реактивность* — моментальная реакция на изменение данных.

8. *Понимаемость* — продукции являются достаточно крупными единицами, интуитивно понятными человеку.

9. *Расширяемость* — продукции могут добавляться в базу знаний или модифицироваться в течение длительного времени без изменения структуры базы знаний. Расширяемость является следствием модульности и декларативности.

К недостаткам продукционного способа представления знаний относят следующие:

1. Процесс вывода *менее эффективен*, чем в других (традиционных) программных системах, поскольку большая часть времени при выводе затрачивается на проверку применимости правила. Однако развитие продукционных систем и увеличение мощности (производительности) компьютеров довольно быстро нивелируют этот недостаток.

2. *Ограниченные возможности контроля правильности* законченной продукционной системы, так как контроль должен осуществляется на самом высшем уровне представления знаний.

3. Родовидовая иерархия понятий *реализуется с большими затруднениями*.

НЕ-фактор

НЕ-фактором называется некоторое понятие, которое лексически, синтаксически и семантически отрицает какое-либо свойство или аспект знания, как, например, противоречивость (отрицает непротиворечивость знания), неточность (отрицает точность знания) и т. д. [21, 22].

Классификация НЕ-факторов по [12]:

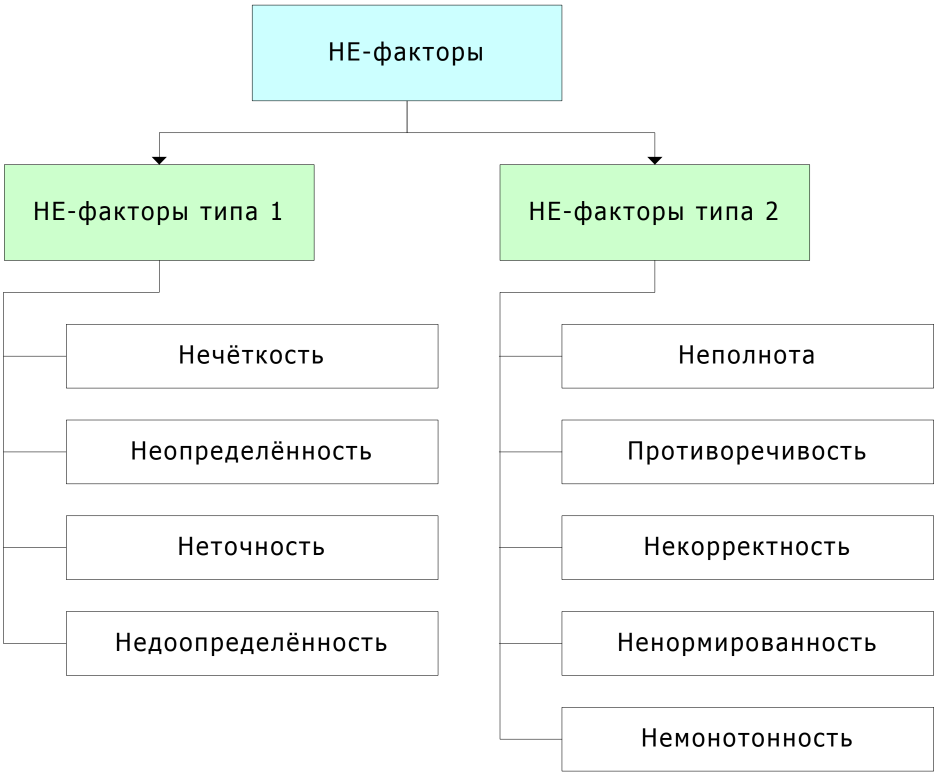

Рисунок 2. Обобщённая классификация НЕ-факторов

Приведённая классификация разбивает всё множество НЕ-факторов на два основных класса — *НЕ-факторы типа 1* и *НЕ-факторы типа 2*. Это разбиение проведено с точки зрения возможности приобретения знаний с НЕ-факторами в автоматизированном режиме из эксперта. Так НЕ-факторы типа 1 можно извлекать из эксперта в автоматизированном режиме при помощи определённых эвристических механизмов [12]. С другой стороны, НЕ-факторы типа 2 не подлежат извлечению из источников знаний первого рода (экспертов), так как они вообще должны быть по возможности устранены из систем, основанных на знаниях. Для устранения проявлений НЕ-факторов типа 2 также используются различные технологии, например для обнаружения неполноты можно использовать механизмы Data Mining [43].

Следующие четыре термина — это определения НЕ-факторов типа 1. Определений НЕ-факторов типа 2 здесь не приводится, так как в дальнейшем изложении они вообще не рассматриваются.

Нечёткость

Аппарат нечёткой логики был разработан в середине XX века Л. А. Заде, как расширение Аристотелевой логики. Нечёткая логика является бесконечнозначной логикой, оперирующей значениями истинности из интервала [0; 1], причём 0 — это полная ложь, 1 — полная истина, а все промежуточные значения интерпретируются в зависимости от решаемой задачи [23].

Используя теоретико-множественный подход можно сказать, что нечёткость проявляется в тех случаях, когда имеет место утверждение $x \in F$, при этом F — нечёткое множество [29].

В общем случае нечёткость в знаниях предполагает, что некоторый параметр x является лингвистической переменной, которая может принимать нечёткие значения [18].

Неопределённость

В соответствии с работой [12] под неопределённостью понимается случай, когда к значениям некоторых параметров проблемной области эксперт приписывает некоторую степень уверенности (которая, в свою очередь, может быть сложной природы). Например, эксперт может явно описывать свою уверенность в высказываниях о проблемной области в виде числа из интервала [0; 1], либо оценивать уверенность интервалом $[a; b] \subseteq [0; 1]$ [39, 52].

Иногда под неопределённостью понимается нечёткость [4], то есть для некоторых задач можно утверждать, что неопределённость — это частный вид нечёткости.

Неточность

Неточность — это один из наиболее часто встречающихся НЕ-факторов [12], так как он проявляется в знаниях тогда, когда при извлечении оцениваются некоторые параметры, полученные при помощи измерительных приборов, которые имеют свою погрешность измерения. Именно погрешность измерения обуславливает то, что измеренные параметры неточны.

В терминах «$x \in F$» неточность определяется как наличие некоторого множества X, имеющего непустое пересечение с F, и при этом значение параметра x определено с точностью до X [34].

Недоопределённость

Недоопределённость — это частичное отсутствие знаний о значении какого-либо параметра (измеримого или нет) [21]. В случае измеримых параметров недоопределённость и неточность можно легко приводить друг к другу, однако существует чёткое разграничение. В случае недоопределённости частичное отсутствие знаний можно восполнять, постепенно доопределяя параметр, а неточные измеренные параметры самодостаточны сами по себе, так как зачастую повышать точность измерения для решения конкретной задачи не имеет смысла (например, бессмысленна точность в один метр при решении гипотетической задачи о перемещении планет в околосолнечном пространстве).

Лингвистическая переменная

Лингвистической переменной называется набор [18]:

$$LV = \langle \beta, T, X, G, M \rangle,$$

где:

β — Наименование лингвистической переменной.

T — Множество её значений (терм-множество), представляющих собой наименования нечётких переменных, областью определения каждой из которых является множество X. Множество T называется базовым терм-множеством лингвистической переменной.

G — Синтаксическая процедура, позволяющая оперировать элементами терм-множества T, в частности, генерировать новые термы (значения). Множество $T \cap G(T)$, где $G(T)$ — множество сгенерированных термов, называется расширенным терм-множеством лингвистической переменной.

M — Семантическая процедура, позволяющая превратить каждое новое значение лингвистической переменной, образуемое процедурой G, в нечёткую переменную, то есть сформировать соответствующее нечёткое множество.

Нечёткая переменная

Нечёткая переменная характеризуется тройкой [18]:

$$FV = \langle \alpha, X, A \rangle,$$

где:

α — Наименование переменной.

X — Универсальное множество (область определения α).

A — Нечёткое множество на X, описывающее ограничения (то есть $\mu_A(x)$) на значение нечёткой переменной α.

Функция принадлежности

В широком смысле под функцией принадлежности понимается характеристическая функция нечёткого множества, которая принимает значения из некоторого упорядоченного множества нечётких оценок [18]. В более узком смысле [23] функция принадлежности нечёткого множества A — это такая функция μ_A, определённая на элементах универсума, которая каждому элементу $x \in U$ (U — универсум, или универсальное множество) ставит в соответствие нечёткую оценку принадлежности x множеству A, при этом обычно считается, что:

$$\mu_A : U \to [0;1].$$

Альфа-срез нечёткого множества (α-срез)

Пусть A — нечёткое множество на области определения X, и есть некоторое число $\alpha \in (0;1]$. Альфа-срезом множества A называется такое множество, что:

$$A_\alpha = \{x \in X \mid \mu_A(x) \geq \alpha\},$$

где $\mu_A(x)$ — функция принадлежности множества A. Нижняя и верхняя граница интервала α-среза равны $\inf_{x \in X} A_\alpha$ и $\sup_{x \in X} A_\alpha$ соответственно.

Степень неопределённости

Степень неопределённости — это обобщённая мера неопределённости какого-либо явления, события, факта [9]. Обычно такой мерой является либо нечёткое значение истинности (в смысле нечёткой логики Л. А. Заде), либо интервальная вероятностная мера, обрабатываемая в рамках теорий вероятности и теории неопределённости Демпстера-Шейфера [39, 52].

Глава 1. Приобретение и извлечение знаний с НЕ-факторами

Как уже говорилось (см. Глоссарий) под приобретением знаний понимается процесс получения знаний от эксперта или каких-либо других источников и формализация этих знаний для последующего использования их в системах, основанных на знаниях.

Необходимо чётко различать термины «приобретение знаний» и «извлечение знаний». В представленной монографии под первым будет пониматься процесс получения знаний из источников знаний при помощи использования программных средств поддержки деятельности инженера по знаниям и эксперта. Под вторым термином будет пониматься получение знаний из экспертов или других источников знаний без использования компьютерных средств поддержки этого процесса, а путём непосредственного контакта инженера по знаниям и источника знаний.

В общем случае, знания Z одного эксперта или некоторой группы экспертов, заносимые в базы знаний интеллектуальных систем, можно представить в следующем виде [36]:

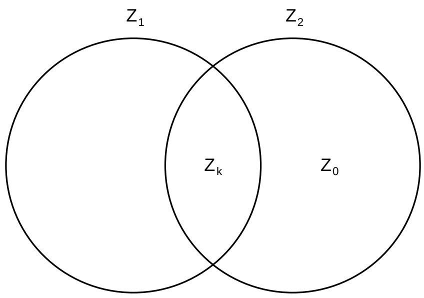

Рисунок 3. Структура знания группы экспертов

где:

- Z_0 — Это чистые эмпирические знания, не подтверждённые теорией. Для различных наук соотношение этой области знания с областью Z_k носит различный характер. Для гуманитарных наук с их «мягкими» знаниями (преимущественно феноменологическими и качественными) характерно преобладание Z_0. В естественных и точных науках с их «жёсткими» знаниями (уровень «количественной» теории) очевидно преобладание Z_k.
- Z_1 — В основе этих знаний лежит теория. Это некоторая идеализация предметной области, но, следовательно, и её упрощение.
- Z_2 — В основе этих знаний лежит опыт. Это более гибкая и широкая часть описательных знаний экспертов, поэтому эти знания не так системны, как Z_1, но они не так искусственны.
- Z_k — Канонизированная часть личных знаний, то есть то, что усвоено экспертами из различных источников (специальной литературы), и в чём нет расхождения между различными экспертами.

В то время как знания Z_1 можно почерпнуть из источников знания второго рода (справочников, учебников и т. п.), знания Z_2 являются «личной интеллектуальной собственностью» каждого конкретного эксперта, при этом не каждый эксперт будет делиться этими знаниями.

Задачей инженерии знаний является получение и формализация в пригодный для обработке вид знаний Z_0.

Так как история нечёткой математики ведёт своё начало с 50-ых годов XX столетия, за прошедшее время было создано значительно количество методов как извлечения, так и приобретения знаний с элементами нечёткости. Поэтому в дальнейшем изложении по возможности будут описаны как методы приобретения, так и методы извлечения нечётких знаний.

Для трёх других выделенных НЕ-факторов: неопределённости, неточности и недоопределённости сложно говорить о применении компьютерных методов поддержки получения знаний из источников, поэтому в соответствующих разделах речь будет вестись исключительно о методах извлечения знаний с этими НЕ-факторами.

1.1. Приобретение и извлечение нечёткости

В основании любой теории из любой области естествознания лежит основополагающе понятие, необходимое для построения самой теории. Для нечёткой математики таким понятием является нечёткое множество, которое характеризуется своей функцией принадлежности [23]. Посредством нечётких множеств можно строго описывать присущие для мышления человека расплывчатые элементы, «без формализации которых нет надежды существенно продвинуться вперёд в моделировании интеллектуальных процессов» [5].

Основной трудностью, мешающей интенсивному применению теории нечётких множеств при решении практических задач, является то, что функция принадлежности должна быть построена вне самой теории и, следовательно, её адекватность не может быть проверена непосредственно средствами теории. В каждом в настоящее время известном методе построения функций принадлежности формулируются свои требования и обоснования к выбору именно такого построения.

Традиционно выделяются две группы методов построения функций принадлежности нечётких множеств: прямые и косвенные методы.

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

- $\forall u_1, u_2 \in U : \mu_A(u_1) < \mu_A(u_2)$ тогда и только тогда, когда u_2 предпочтительнее u_1, то есть в большей степени характеризуется понятием A;
- $\forall u_1, u_2 \in U : \mu_A(u_1) = \mu_A(u_2)$ тогда и только тогда, когда u_1 и u_2 безразличны относительно понятия A.

Примеры прямых методов: непосредственное описание функции принадлежности в виде таблицы, формулы, примера.

В косвенных методах значения функций принадлежности выбираются таким образом, чтобы удовлетворить заранее сформулированным условиям. Экспертная информация является только исходной информацией для дальнейшей обработки. Дополнительные условия могут налагаться как на вид получаемой информации, так и на процедуру обработки. Примерами дополнительных условий могут служить следующие: функция принадлежности должна отражать близость к заранее выделенному эталону, объекты множества U являются точками в параметрическом некотором пространстве [53]; результатом процедуры обработки должна быть функция принадлежности, удовлетворяющая условиям интервальной шкалы [13]; при попарном сравнении объектов, если один объект оценивается в α раз сильнее, чем другой, то второй объект оценивается только в $1/\alpha$ раз сильнее, чем первый [50].

Как правило, прямые методы используются для описания понятий, которые характеризуются измеримыми свойствами, как-то высота, вес, объём, рост, время. В этом случае удобно непосредственное задание значений степеней принадлежности. К прямым методам можно отнести те методы, которые основаны на вероятностной трактовке функции принадлежности [23]: $\mu_A(u) = P(A \mid u)$, то есть вероятность того, что объект $u \in U$ будет отнесён к множеству, которое характеризует понятие A.

Примерная классификация методов построения функций принадлежности по [23] приведена на следующем рисунке:

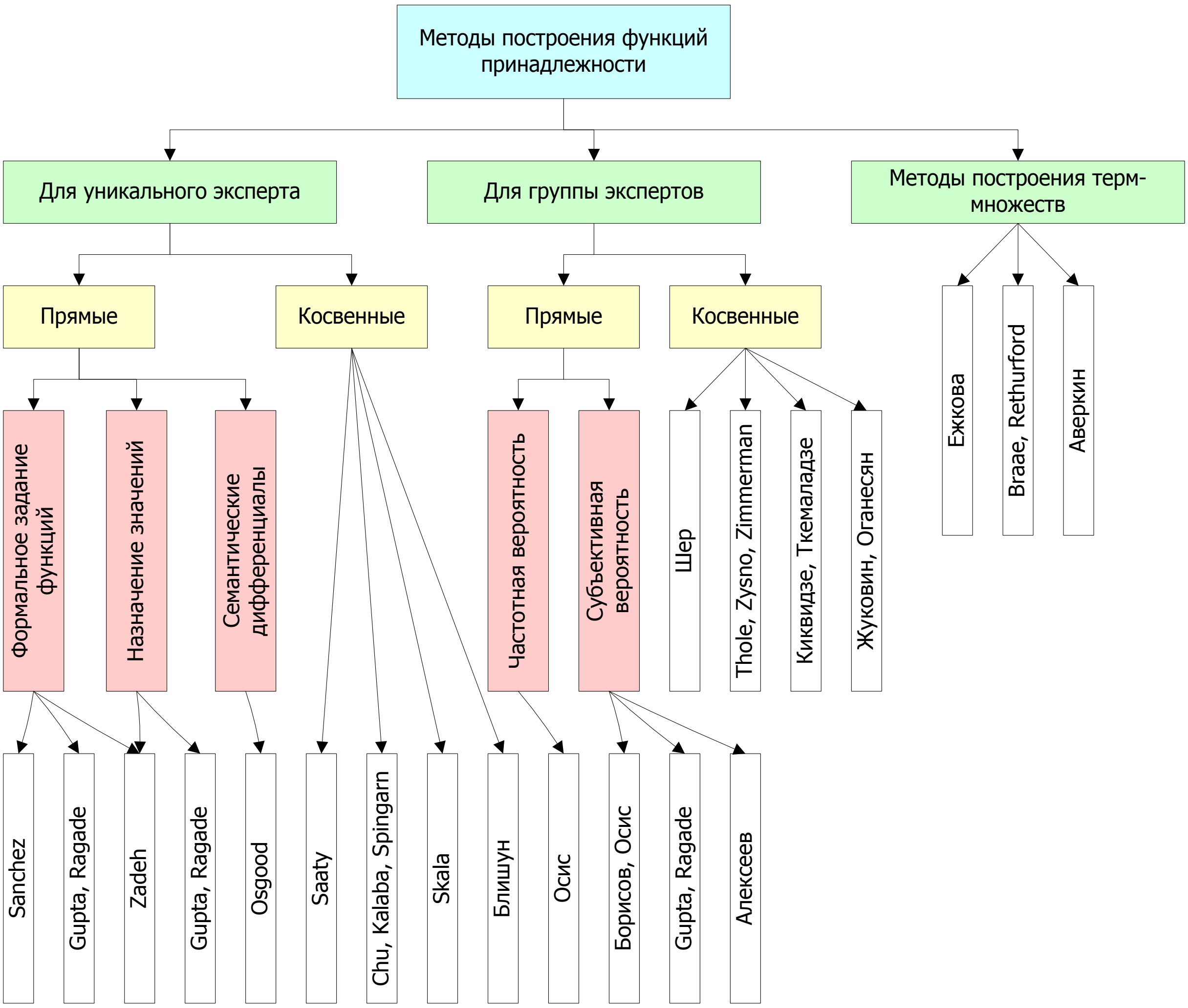

Рисунок 4. Классификация методов построения функций принадлежности

В мышлении человека порядок создаётся из хаоса путём формирования системы полярных шкал и различения некоторых объектов с помощью оценок на этих шкалах. Концевые значения оппозиционных шкал соответствуют некоторым противоположным свойствам, выражаемым парами антонимов, например, «активный — пассивный», «тупой — острый», «добрый — злой» [31].

Если предполагается, что люди далеки от случайных ошибок и работают как «надёжные и правильные приборы», то можно спрашивать эксперта непосредственно о значениях принадлежности объектов на той или иной шкале. Однако имеются искажения [54], например, субъективная тенденция сдвигать оценки объектов в направлении концов оценочной шкалы. Следовательно, прямые измерения, основанные на непосредственном определении принадлежности, должны использоваться только в том случае, когда такие ошибки незначительны или маловероятны.

Функции принадлежности могут отражать мнение как некоторой группы экспертов, так и одного уникального эксперта. Комбинируя возможные два метода построения функций принадлежности (прямой и косвенный) с одним или несколькими экспертами, можно получить четыре типа экспертизы [6]. Все четыре типа экспертизы рассматриваются в следующих разделах.

1.1.1. Прямые методы для одного эксперта

Прямые методы для одного (уникального) эксперта состоят в непосредственном назначении степени принадлежности для исследуемых объектов или непосредственном назначении функции (правила), позволяющей вычислять значения. Для примера можно рассмотреть построение функции принадлежности понятия «старый» в отношении человека. Пусть эксперт руководствуется следующими рассуждениями: переменная «возраст» принимает значения из интервала $U = [0, 100]$. Слово «старый» можно интерпретировать как имя нечёткого подмножества U, которое характеризуется некоторой функцией совместимости. Таким образом, степень, с которой численное значение возраста, например $u = 72$, совместимо с понятием «старый» есть 0.9, в то время как совместимость 70 и 65 с тем же понятием есть 0.85 и 0.8 соответственно. Эквивалентно функция $\mu_{старый}(u)$ может рассматриваться как функция принадлежности нечёткого множества «старый».

В [49] предложен метод семантических дифференциалов. Практически в любой области можно выделить множество шкал оценок, используя следующую процедуру:

1. Определить список свойств, по которым оценивается понятие (объект).
2. Найти в этом списке полярные свойства и сформировать полярные шкалы.
3. Для каждой пары полюсов оценить исследуемое понятие на то, как сильно оно обладает положительным свойством (можно использовать для оценки числа от -3 до 3 или от 1 до 7, а также интервалы от 0 до 100 %, либо от 0 до 10).

Совокупность оценок по шкалам называется «*профилем понятия*». Следовательно, вектор с координатами, изменяющимися от 0 до 1, также называется профилем. Профиль есть нечёткое подмножество положительного списка свойств или шкал.

Например, в задаче распознавания лиц можно выделить следующие 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: тёмные / светлые
X_6 — форма подбородка: остроконечный / квадратный
X_7 — толщина губ: тонкие / толстые
X_8 — цвет лица: тёмное (смуглое) / светлое (белое)
X_9 — очертание лица: овальное / квадратное

Светлое, квадратное лицо, у которого чрезвычайно широкий лоб, курносый длинный нос, широкие и светлые глаза, остроконечный подбородок, может быть определено как нечёткое множество $\{1|X_1, ..., 1|X_9\}$ или вектор (111 111 111). Лицо, соответствующее вектору (000 000 000), полярно противоположно.

При вычислении частичной принадлежности строгих множеств друг другу можно использовать следующий метод. Пусть покрытием K обычного множества U является любая совокупность обычных подмножеств $\{A_1, ..., A_k\}$ множества U: $A_1 \cup ... \cup A_k = U$. В крайнем случае, когда для любых i и j таких, что $i \neq j$, $A_i \cap A_j = \varnothing$, имеет место разбиение U. Пусть имеется $B \subseteq U$, тогда B может рассматриваться как нечёткое подмножество K с функцией принадлежности, вычисляемой по формуле:

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

где |A| — мощность множества A.

В качестве примера можно рассмотреть несколько абстрактную ситуацию, описываемую параметрами:

$U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

$K = \{\{1, 3, 5\}, \{3, 6, 9\}, \{2, 4, 8\}, \{1, 3, 7\}, \{2, 3, 8\}\} = \{A_1, A_2, A_3, A_4, A_5\}$

$B = \{2, 3, 5, 9, 8\}$

Тогда если рассматривать B как нечёткое подмножество K, по вышеприведённой формуле можно получить: $B = \{1/3 \mid A_1, 1/3 \mid A_2, 1/3 \mid A_3, 1/7 \mid A_4, 3/5 \mid A_5\}$, или как набор значений частичной принадлежности $\mu_B = \{1/3, 1/3, 1/3, 1/7, 3/5\}$.

Этот метод можно применять в случае, если нечёткие множества необходимо получить из дискретных строгих множеств, мощность которых исчислима и не слишком велика. Например, метод можно использовать в ставшем уже классическом примере (в различных вариациях) получения функций принадлежности для названий машин в строгом множестве самих машин. Кроме того, этот метод также можно применять для *фаззификации* чётких величин, значения которых входят в некоторые строгие множества.

Ещё один метод основан на том, что любое решение задачи многоцелевой оптимизации рассматривается как нечёткое подмножество значений целевых функций следующим образом. Пусть $f_1, ..., f_r$ — целевые функции, где $f : \mathbf{R}^n \to \mathbf{R}$, и пусть требуется решить задачу $f_i \to \max$ для всех i. Пусть $f_i^* < \infty$ — максимальное значение функции f_i, которое не зависит от других функций, а $C = \{f_1, ..., f_r\}$ — множество целевых функций, тогда любое значение x в области определения f_i такое, что $f_i(x) \le f_i^*$, можно рассматривать как нечёткое множество на C с вектором значений принадлежности $\mu_x = (\mu_1, ..., \mu_r)$, где μ_i вычисляется по формуле:

$$\mu_i = \frac{f_i^* - f_i(x)}{f_i^*}.$$

1.1.1.1. Шкалы для определения экспертных оценок

В работе [4] рассматривается использование порядковых шкал в задачах экспертной классификации. Предполагается, что диагностируемые свойства объектов исследования могут иметь различную степень выраженности и быть естественным образом упорядоченными. Используется гипотеза о различной степени характерности отдельных значений каждого признака для каждого свойства. Предполагается, что по каждому признаку эксперт может упорядочить его значения по их характерности для соответствующего класса, и этот порядок не зависит от значений других признаков. Эти классы могут отражать различные степени уверенности эксперта в наличии или отсутствии некоторого свойства, либо степень выраженности этого свойства (или то и другое вместе). Степень уверенности эксперта в своём ответе может выражаться в виде суждений типа: «с большой степенью уверенности можно констатировать наличие некоторого свойства в данном состоянии объекта», «маловероятно, что у объекта в данном состоянии присутствует данное свойство» и т. п. Возможные степени выраженности свойства могут оцениваться как «сильная», «средняя», «слабая». Отмечается существование границ возможностей человека в задачах порядковой классификации, а также ограниченная ёмкость кратковременной памяти человека.

Анализ вербальных шкал, используемых в экспертном оценивании, позволяет выделить некоторую унифицированную структуру подобных шкал, которая может быть выбрана в качестве шкалы для измерения степени уверенности. Различным вербальным шкалам, используемым на практике, присущи следующие черты:

1. Качественность оценок, использование словесных оценок для измерения свойств, для которых не разработаны количественные шкалы. Примеры: «уродливый», «не очень трудолюбивый», «совсем безопасный», «абсолютно пригодный», «слишком бодрый», «слегка помятый».

2. Приблизительность оценок, использование их даже тогда, когда свойство может быть измерено в количественной шкале. Примеры: «молодой», «маленький», «совсем легкий», «достаточно медленный», «очень холодный».

3. Использование противоположных бинарных оценок: «плохой — хороший», «молодой — старый», «сильный — слабый», «опасный — безопасный», «усталый — бодрый», «мягкий — твердый» и т. д.

4. Наличие нейтральной оценки: «ни часто, ни редко», «средних лет», «предпочтение средней силы».

5. Использование пяти — семи градаций при оценке свойств.

Наличие общих черт, присущих различным вербальным шкалам, позволяет выделить общую структуру шкал, по которым измеряются свойства объектов:

$$L = \{eap < vap < ap < np < p < vp < ep\},$$

где: p — базовая градация измеряемого свойства P;

ap (“anti-p”) — базовая градация шкалы, противоположной p;

v (“very”) и e (“extra”) — модификаторы базовых градаций;

np — нейтральная, средняя градация шкалы.

Эта шкала была названа унифицированной шкалой или шкалой с унифицированной структурой. Такая унифицированная структура широко применяется при построении нечётких шкал, градациями которых являются лингвистические термы — нечёткие величины, рассматриваемые обычно как значения нечетких переменных. Очень часто такая унифицированная шкала используется при построении отношений моделирования для некоторых измеряемых свойств P (см. далее).

1.1.1.2. Классификация шкал

Как показано в предыдущем разделе шкалы имеют довольно весомую роль в процессах приобретения знаний с НЕ-факторами, особенно с нечёткостью. Для конкретизации рассмотрения вопросов, связанных с применением шкал, необходимо привести более или менее обобщённую классификацию типов шкал, для чего можно воспользоваться в частности работой [17].

На следующем рисунке показана обобщённая классификация типов шкал:

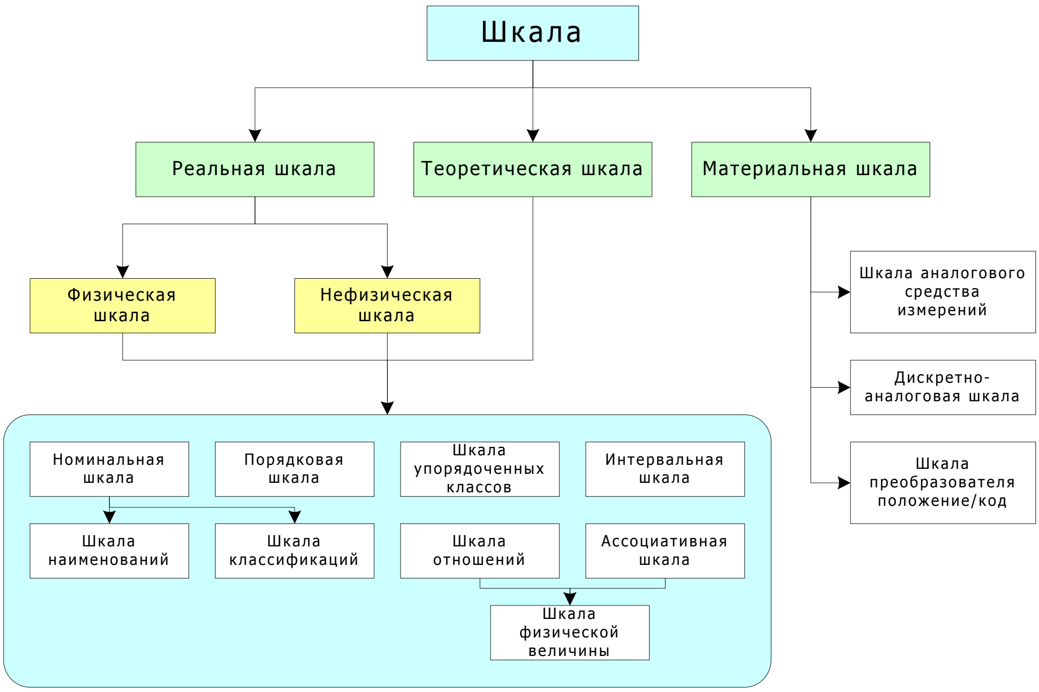

Рисунок 5. Обобщённая классификация шкал

Далее приведены определения различных типов шкал.

Шкала

Совокупность системы объектов, системы чисел (или знаков) и правил, позволяющих адекватно отобразить систему объектов в систему чисел (знаков).

1. В соответствии с этим определением шкалу можно образно представить как своего рода словарь, делающий возможным перевод с «языка вещей» на язык знаков (и обратно) при выполнении когнитивных процедур.
2. Система объектов, которая, согласно этому определению, входит в состав шкалы, может быть потенциально бесконечной. В подобных случаях включение системы объектов в состав шкалы следует понимать, как возможность указать при соответствующих условиях для любого возможного объекта его место на шкале.
3. Знаки, входящие в состав шкалы, могут быть числовыми знаками (последовательностями цифр), лингвистическими знаками, кодовыми знаками (например, электрическими импульсами) и т. д.
4. Адекватность отображения понимается в том смысле, что правила присваивания чисел (знаков) объектам обеспечивают соответствие определённых, свойственных рассматриваемой шкале, отношений в системе объектов отношениям, принятым для системы чисел или знаков.
5. Это определение относится к шкале в общем смысле слова. Наряду с этим допускается использование термина «шкала» в качестве краткой формы составных терминов «шкала аналогового средства измерений», «шкала преобразователя положение/код» и т. п., если контекст исключает неверное понимание.
6. Следует избегать употребления термина «шкала» как синонима термина «диапазон» (например, в сочетаниях типа «конец шкалы»). Вместе с тем, оправдано выражение «длина шкалы цифрового прибора» — число ступеней квантования в его диапазоне.

Теоретическая шкала

Шкала, в которой в качестве объектов выступают идеальные модели реальных объектов, и в системе этих моделей постулируются определённые отношения и операции, отражаемые отношениями и операциями в системе чисел (или лингвистических знаков).

Реальная шкала

Шкала, в которой в качестве объектов выступают реальные объекты (физические тела, системы тел, поля, материальные процессы и их состояния, а также, возможно, состояния психических, социальных, экономических, культурных и т. п. явлений и процессов), в системе которых поддаются выявлению определённые отношения и могут быть выполнены операции, отражаемые отношениями и операциями в системе чисел (знаков).

Материальная шкала

Совокупность хранимых или воспроизводимых материальных объектов (тел, состояний процессов и т. п.), помеченных цифровыми или иными знаками. Например, шкала аналогового средства измерений, шкала преобразователя положение/код, шкала меток времени цифрового частотометра.

Физическая шкала

Реальная шкала, отношения в системе объектов которой поддаются выявлению объективными экспериментальными методами. Определение «физическая» не следует понимать как указание на принадлежность науке физике: физические шкалы используются не только в физике, но и в химии, биологии и других областях знания, а также в практической деятельности людей.

Нефизическая шкала

Реальная шкала, отношения в системе объектов которой поддаются выявлению путём опроса людей или анализа текстов.

Номинальная шкала

Шкала, допустимыми преобразованиями которой являются преобразования группы перестановок.

Шкала наименований

Номинальная шкала, при построении которой учитывается только отношение тождественности каждого объекта самому себе.

Шкала классификаций

Номинальная шкала, при построении которой учитывается отношение эквивалентности объектов в каком-либо аспекте.

Порядковая шкала

Шкала, при построении которой учитываются отношения эквивалентности и порядка объектов в каком-либо аспекте, а допустимыми преобразованиями являются положительные монотонные преобразования.

Шкала упорядоченных классов

Шкала, разбивающая множество объектов на ряд классов так, что между объектами различных классов имеется отношение порядка.

Интервальная шкала

Шкала, допустимыми преобразованиями которой являются преобразования общей линейной группы: $y = ax + b$.

Шкала отношений

Шкала, допустимыми преобразованиями которой являются преобразования подобия: $y = ax$.

Ассоциативная шкала

Шкала, основанная на предположении о том, что шкальное свойство рассматриваемых объектов связано стабильной монотонной зависимостью с другим свойством хотя бы некоторой части этих объектов, причём для этого другого свойства построена шкала отношений. Под шкальным свойством понимается свойство системы объектов, проявляющееся в отношениях, постулируемых или выявляемых при построении и использовании некоторой конкретной шкалы.

Шкала физической величины

Реальная общезначимая ассоциативная шкала или шкала отношений, построенная для конкретной физической величины. Шкала называется общезначимой, если для этой шкалы обеспечена возможность указать с известной степенью неопределённости места любых объектов, имеющихся в наличии в разное время и в разных местах.

Шкала аналогового средства измерений

Поверхность с видимыми глазом отметками (штрихами или иными знаками малой протяжённости) и, возможно, числовыми и другими знаками, предназначенная для отсчёта значения измеряемой или воспроизводимой величины по положению указателя, перемещающегося относительно отметок.

Дискретно-аналоговая шкала

Шкала аналогового средства измерений, снабжённая совокупностью последовательно расположенных элементов, один из которых или некоторая их часть может светиться или иным образом изменять вид, создавая для глаза впечатление перемещающегося указателя.

Шкала преобразователя положение/код

Устройство в виде диска, барабана или линейки, различные участки которого различаются по физическим свойствам, что позволяет с помощью соответствующих воспринимающих элементов формировать кодовые сигналы, зависящие от положения воспринимающих элементов относительно шкалы.

Таким образом, в большинстве случаев приобретения нечётких знаний будет использоваться некоторая конкретная шкала физической величины. Именно на такой шкале должны строиться функции принадлежности. При этом название физической величины, для которой построена шкала, будет в то же время являться названием лингвистической переменной, для которой строятся функции принадлежности терм-множеств.

1.1.2. Косвенные методы для одного эксперта

На практике часто имеют место случаи, когда не существует элементарных измеримых свойств или признаков, через которые определяются рассматриваемые понятия, например, «красота», «интеллект». В таких случаях вызывает затруднение задача ранжирования степени проявления свойства у рассматриваемых элементов (объектов). Так как степени принадлежности рассматриваются на определённом множестве объектов, а не в абсолютном смысле, то интенсивность принадлежности можно определить исходя из попарных сравнений рассматриваемых элементов. Если бы значения степеней принадлежности были известны, то попарные сравнения можно представить матрицей отношений $A = ((a_{ij}))$, где $a_{ij} = \omega_i/\omega_j$, $\omega_i = \mu_S(u_i)$, S — рассматриваемый элемент.

Если отношения точны, то получается соотношение $A\omega = n\omega$, где n — собственное значение матрицы A, $\omega = (\omega_1, ..., \omega_n)$, по которому можно восстановить вектор ω, с учётом того, что вектор нормализован:

$$\sum_{i=1}^{k} \omega_i = 1.$$

Так как отношения сравнения a_{ij} в реальных случаях неточны из-за того, что они получены эмпирическим способом, необходимо вычислить оценки для ω. Для улучшения согласованности оценок в рассматриваемом методе предполагается, что $a_{ij} \cdot a_{jk} = a_{ik}$, откуда следует, что все диагональные элементы матрицы A равны 1, а для симметричных относительно главной диагонали элементов истинно соотношение: $a_{ij} = 1/a_{ji}$. Грубо говоря, если элемент оценивается в α раз сильнее другого, то другой элемент оценивается всего в 1/α раз сильнее первого. Если имеется полная согласованность в рассуждениях эксперта относительно степеней оценки a_{ij} (согласованности по транзитивности), то ранг матрицы A равен 1, и чтобы решить поставленную задачу, достаточно знать элементы только по одну сторону главной диагонали A.

В этом случае

$$\sum_{j=1}^{n} a_{ij} \frac{\omega_j}{\omega_i} = n, \; i = \overline{1, n},$$

где n — наибольшее собственное значение матрицы A, а другие собственные значения λ равны нулю, так как сумма всех собственных значений должна равняться n. В общем случае эмпирическая шкала $\omega = (\omega_1, ..., \omega_n)$ должна удовлетворять задаче на поиск собственного значения $A\omega = \lambda_{max}$, где λ_{max} — наибольшее собственное значение. Чем ближе λ_{max} к числу n, тем более верным является результат. Отклонение λ_{max} от n используется как мера полезности (правильности) результата. В процедуре решения задачи формируется матрица сравнений рассматриваемого множества элементов. Элементы такой матрицы — это значения, показывающие во сколько раз один элемент лучше другого. Так как известно, что задача $A\omega = \lambda_{max}\omega$ имеет единственное решение, то значение координат собственного вектора, соответствующего максимальному собственному значению, делённые на их сумму, будут искомыми степенями принадлежности.

При формировании оценок попарных сравнений, обычно эксперта просят отразить ощущения или опыт следующим образом: а) установить, какой из двух предлагаемых элементов на его взгляд более важен; б) оценить восприятие интенсивности различия в виде ранга важности по определённой ранговой шкале.

В следующей таблице приводятся качественные оценки и соответствующие им численные значения, обычно используемые в подобных методах:

Таблица 2. Качественные оценки, используемые в методах парных сравнений

Интенсивность важности	Качественная оценка	Объяснения
0	Несравнимость	Нет смысла сравнивать элементы
1	Одинаковая значимость	Элементы равны по значимости
3	Слабо значимее	Существуют показания о предпочтении одного элемента над другим, но показания неубедительные
5	Сильно значимее	Существует хорошее доказательство и логические критерии, которые могут показать, что элемент более важен
7	Очевидно значимее	Существует убедительное доказательство большей значимости одного элемента над другим
9	Абсолютно значимее	Максимально подтверждается ощутимость предпочтения одного элемента над другим
2, 4, 6, 8	Промежуточные оценки между соседними значениями	Используются, когда необходим компромисс
Обратные значения для ненулевых оце-	Если оценка a_{ij} имеет ненулевое значение, приписан-	

нок	ное на основании сравнения элемента i с элементом j, то $a_{ji} = 1/a_{ij}$.	
Нормирование	Нормирование возникает из описанной шкалы	

Предполагается, что элементы с нулевой оценкой не рассматриваются при попарном сравнении. При анализе сложных свойств, которые представляются как иерархическая система, предлагается использовать описанный метод при сравнении составляющих свойств на удовлетворение (соответствие) сложному свойству.

В случае если в матрице парных сравнений отсутствуют некоторые элементы, можно воспользоваться методом, предлагаемом в [6]. Рассматривается понятие «класс S», которое описывается функцией принадлежности на множестве объектов $A = \{a^0, ..., a^{n-1}\}$. В A имеется только два объекта, о которых можно сказать, что a^1 — идеальный представитель тех объектов, которые принадлежат S, и что a^0 — идеальный представитель тех объектов, которые не принадлежат понятию «класс S», то есть $\mu_S(a^1) = 1$, $\mu_S(a^0) = 0$. Эксперту предлагается проранжировать степень различия объектов в каждой паре объектов в смысле принадлежности понятия классу S. В результате формируется матрица попарных сравнений, которая задаёт порядок пар объектов по степени различия в парах. Далее посредством методов неметрического шкалирования [16] в факторном (метрическом) пространстве X^m вычисляются координаты n точек $x^i = \{x^i_1, ..., x^i_m\}$, порядок расстояний $d(x^i, x^j)$ между которыми совпадает или максимально близок к порядку элементов матрицы попарных сравнений. Для полученных расстояний имеют место следующие утверждения:

- Если объекты a^i и a^j неразличимы, то $d_{ij} = 0$.
- Если степень различия объектов a^i и a^j больше, чем степень различия объектов a^i и a^k, то $d_{ij} > d_{ik}$.
- Если степень различия объектов a^i и a^j совпадает со степенью различия объектов a^i и a^k, то $d_{ij} = d_{ik}$.

Дальнейшие выводы основываются на следующих двух предположениях:

- Предположение 1. Понятие S характеризуется несколькими одномерными признаками, которые определяются при помощи методов неметрического шкалирования.

- Предположение 2. Степень различия двух объектов a^i и a^j из A по отношению к понятию S пропорциональна разности расстояний в пространстве признаков от a^i и a^j до объекта a^1, который с максимально возможной степенью принадлежит понятию S.

Согласно первому предположению объекты формально описываются точками в пространстве признаков. Наличие нескольких признаков позволяет объяснить, например, нетранзитивность в парных сравнениях. Из процедуры получения формального описания объектов следует, что максимальное расстояние на множестве объектов будет между объектами a^0 и a^1, так как их различие в смысле принадлежности понятию S будет максимально возможным. Следовательно, чем дальше исследуемый объект a^i от эталона a^1 в пространстве признаков, тем в меньшей степени он характеризуется понятием S.

Из второго предположения следует, что степень различия двух объектов a^i и a^j по отношению к понятию S будет пропорциональна разности значений функции принадлежности на этих объектах, то есть имеет место равенство:

$$c\left|d_{1i} - d_{1j}\right| = \left|\mu_S(a^i) - \mu_S(a^j)\right|,$$

где c — некоторая константа. Если в качестве объекта a^i последовательно рассматривать объекты a^0 и a^1, то можно вывести следующие закономерности:

$$\begin{array}{c} c(d_{10} - d_{1j}) = \mu_S(a^j) \\ cd_{1j} = 1 - \mu_S(a^j) \end{array}.$$

Из этих уравнений следует, что

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

Таким образом, функция принадлежности на множестве объектов A, характеризующая понятие S, определяется по расстояниям в пространстве признаков X^m согласно полученного соотношения.

1.1.3. Прямые методы для группы экспертов

При интерпретации степени принадлежности как вероятности возникает два момента. Вероятность можно понимать как объективную и как субъективную. В случае объективного понимания вероятности можно вычислять функции принадлежности для нескольких классов понятий S_i при помощи использования равенства $\mu_{Sj}(u_i) = p(S_j|u_i)$, где условная вероятность определяется по формуле Байеса:

$$p(S_j \mid u_i) = \frac{p_{u_i}(S_j)p(u_i \mid S_j)}{\sum_{j=1}^{m} p_{u_i}(S_j)p(u_i \mid S_j)},$$

причём

$$p_{u_i}(S_j) = \frac{(y_j)_{u=u_i}}{n}, \quad j = \overline{1,m}, \quad i = \overline{1,n},$$

где y_j — число случаев при значении параметра u_i, когда верной оказалась j-ая гипотеза.

Если рассматривать степень принадлежности в качестве субъективной вероятности (то есть вероятности того, что лицо, принимающее решения, отнесёт элемент $u \in U$ к множеству S), то в качестве значения функции принадлежности $\mu_S(u)$ берётся как раз субъективная вероятность того, что эксперт использует понятие S (некоторое понятие из естественного языка) в качестве имени объекта u. При этом считается, что множество U является экстенсионалом понятия S.

Оценивать функции принадлежности можно иначе. Первоначально необходимо определить то максимальное количество классов, которое может быть описано рассматриваемым набором параметров. Для каждого элемента u значение функции принадлежности класса S_1

дополняет до единицы значение функции принадлежности класса S_2 (в случае двух классов). Таким образом, система классов должна состоять из классов, представляющих противоположные события. Сумма значений функций принадлежности произвольного элемента u к системе таких классов всегда должна равняться единице. Если же число классов и их состав чётко не определены, то необходимо вводить условный класс, включающий те классы, которые не выявлены. После этого в процентах оценивается степень проявления каждого класса из заданного перечня в данном состоянии u.

Однако в некоторых случаях мнение эксперта очень сложно выразить в процентах, поэтому более приемлемым способом оценки функции принадлежности является метод опроса, состоящий в следующем. Оцениваемое состояние предъявляется большому числу экспертов. Каждый эксперт имеет один голос. Он должен однозначно отдать предпочтение одному из классов из заранее известного перечня. Значение функции принадлежности вычисляется по формуле:

$$\mu_S(u) = \frac{n_S}{n},$$

где n — число экспертов, участвовавших в эксперименте, n_S — число экспертов, проголосовавших за класс S. Пример из [23]: пусть в результате переписи населения в некоторой области численностью p получено множество значений возраста U от 0 до 100 лет. Пусть y (u) — число людей, имеющих возраст u и утверждающих, что являются молодыми. Пусть n (u) — действительно число людей, имеющих возраст u, тогда:

$$p = \int_0^{100} dn(u).$$

Можно считать, что понятие «молодой» описывается нечётким множеством на U с функцией принадлежности μ (u) = y (u) / n (u). Очевидно, что для малых значений возраста от 0 до 20 лет y (u) = n (u), следовательно μ (u) = 1. Однако не все n (35) считают себя молодыми, следовательно, y (35) < n (35). Для u > 80 число y (u) должно быть очень маленьким.

1.1.4. Косвенные методы для группы экспертов

В работе [35] предлагается способ определения функции принадлежности на основе интервальных оценок. Пусть интервал $[x_{ji}, x'_{ji}]$ отражает мнение i-го эксперта, (i = 1, ..., m) и m > 1 о значении j-го (j = 1, ..., n) признака оцениваемого понятия S. Тогда полным описанием этого понятия i-м экспертом является гиперпараллелепипед:

$$\theta_i = \left[x_{1i}, x'_{1i}\right] \times \ldots \times \left[x_{ni}, x'_{ni}\right].$$

Приводится процедура, позволяющая вычислять коэффициенты компетентности экспертов, а также сводить исходную «размытую» функцию (усреднённые экспертные оценки) к характеристической функции неразмытого, чёткого множества. Алгоритм следующий.

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

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

2. Образовать на основе полученных покрытий непересекающиеся гиперпараллелепипеды:

$$T_k = \left[x_{1k}, x'_{1k}\right] \times \ldots \times \left[x_{nk}, x'_{nk}\right],\ k = \overline{1, m'}.$$

3. Вычислить для $x \in T_k$:

$$\varphi_i(x) = \begin{cases} 1,\ \textit{если}\ T_k \cap \theta_i \neq \varnothing. \\ 0,\ \textit{если}\ T_k \cap \theta_i = \varnothing. \end{cases}$$

4. Номер итерации $l = 1$.

5. Вычислить коэффициенты компетентности:

$$\left\{\lambda_i^l\right\}_{i=1}^m = \left\{\frac{1}{m}\right\}_{i=1}^m$$

6. Вычислить приближение функции принадлежности при нормированных:

$$\lambda_i : \sum \lambda_i^l = 1,\quad f^l(x) = \sum_{i=1}^m \varphi_i(x)\lambda_i^l,\quad x \in T_k,\quad k = \overline{1, m'}.$$

7. Вычислить функционал рассогласования мнения i-го эксперта с мнением экспертного совета на l-й итерации:

$$\delta_i^l = \sum_{\substack{x \in T_k \\ k = \overline{1, m'}}} \left[f^l(x) - \varphi_i(x)\right]^2,\ i = \overline{1, m}.$$

8. Вычислить параметр:

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

9. Увеличить итерацию: $l = l + 1$.

10. Вычислять очередное значение коэффициентов компетентности:

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

11. Если величина $\max |\lambda_i^{l-1} - \lambda_i^l|$ близка к нулю, то прекратить вычисления и считать приближением функции принадлежности $f(x) = \mu_S(x)$, в противном случае возвратиться на шаг 6.

Другой косвенный метод для группы экспертов состоит в упорядочивании элементов, для которых строятся функции принадлежности. Пусть U — универсальное множество, S — понятие, общее название элементов (концепт). Задача определения нечёткого подмножества U, описывающего понятие S, решается путём опроса экспертов. Каждый эксперт E_i, $i = 1, \ldots, m$ выделяет из U множество элементов Q_i, по его мнению, соответствующих концепту S. Ранжируя все элементы множества $Q = \bigcup_{i=1}^{m} Q_i$ по предпочтению в смысле соответствия понятию S, каждый эксперт упорядочивает Q, используя отношения порядка (>), (≥) или (≈). Отношение (≈) указывает на одинаковую степень предпочтения между любыми q_α, $q_\beta \in Q$. Предполагается, что эксперты могут ставить коэффициенты степени предпочтения γ перед элементами в упорядоченной последовательности, усиливая или ослабляя отношение предпочтения. Вводится расстояние между элементами указанной последовательности:

$$\rho(q^i_\alpha, q^i_\beta) = \frac{1}{\gamma} \sum_{j=\alpha}^{\gamma=\beta-1} \gamma_{\gamma+1} \rho(q^i_j, q^i_{j+1}),$$

где

$$\rho(q_l, q_{l+1}) = \begin{cases} 1, \text{ если } q_l > q_{l+1} \\ 1/2, \text{ если } q_l \geq q_{l+1} \\ 0, \text{ если } q_l \approx q_{l+1} \end{cases}.$$

Здесь α, β — порядковые номера элементов в упорядочении. Расстояние вычисляется через первый в упорядочении элемент:

$$\rho(q^i_\alpha, q^i_\beta) = \rho(q^i_1, q^i_\beta) - \rho(q^i_1, q^i_\alpha) = \rho^i_\beta - \rho^i_\alpha.$$

Эта разность показывает насколько предпочтительнее q^i_α по сравнению с q^i_β. При решении задачи взвешивания предпочтительности элементов множества Q предполагается, что разность между весами ($\varphi(q^i_\alpha) - \varphi(q^i_\beta)$) пропорциональна разности ($\rho^i_\beta - \rho^i_\alpha$):

$$\varphi(q^i_{\beta+\nu}) - \varphi(q^i_\beta) = C(\rho^i_{\beta+\nu} - \rho^i_\beta)$$

Когда $\nu = 1$, формула превращается в рекуррентную формулу, и задача сводится к определению веса первого элемента. При использовании рекуррентных формул вес последнего элемента должен отличаться от нуля. Например, в качестве $\varphi(q^i_\alpha)$ можно выбрать $\max_\alpha(\rho^i_\alpha - \rho_0)$. На основании всех $\varphi(q^i_\alpha)$, i = 1, ..., m для q_α определяется значение

$$\varphi(q_\alpha) = \frac{1}{m} \sum_{i=1}^{m} \varphi(q^i_\alpha),$$

которое и есть степень принадлежности элемента $u \in U$ нечёткому множеству с общим названием S.

В [13] предлагается следующий метод отображения множества объектов $U = \{u_1, ..., u_n\}$ во множество действительных чисел из [0, 1]. Эксперту предлагаются всевозможные пары объектов из U. В результате эксперимента с ν-м экспертом получается матрица $\|\delta^\nu_{ij}\|$, где i, j = 1, ..., n; а δ^ν_{ij} равно 1 в случае, если эксперт ответил $u_i \geq u_j$ и равно нулю в противном случае. В результате опроса N экспертов будет сформировано N матриц. После этого вводятся новые величины

$$n_{ij} = \sum_{\nu=1}^{N} \delta_{ij}^{\nu},$$

указывающие число голосов, поданных за решение u_j против решения $t_{ij} = n_{ij} - n_{ji} = 2n_{ij} - N$. Значения функции принадлежности определяются следующей формулой:

$$\mu_i = \sum_{i=1}^{n} \lambda_i t_{ij},\ j = \overline{1,n},\ \lambda_j \geq 0,\ \sum_{j=1}^{n} \lambda_j = 1.$$

На основании этого представления каждому решению u_i приписывается число μ_i в интервальной шкале, если выполняются условия:

$$\Delta\mu_{ij} = \mu_i - \mu_j = z_{ij},\ \Delta\mu_{ij} = \Delta\mu_{lj} + \Delta\mu_{il}.$$

Поскольку эксперимент с экспертами протекает произвольно, то следует ожидать, что будет нарушено условие $t_{il} = z_{ij} + z_{jl}$. В работе [13] приводится метод сглаживания, позволяющий получить новые элементы z_{ij}, которые и определяют $\Delta\mu_{ij}$.

1.1.5. Построение отношения моделирования

Для практических задач достаточно наличия нечёткого языка с фиксированным конечным словарём. Это ограничение не слишком сильное, так как лингвистические переменные, используемые при решении задач, имеют базовое терм-множество, состоящее, как правило, из 2 – 9 термов [23]. Каждый терм описывается нечётким подмножеством множества значений U некоторой базовой переменной u и рассматривается как лингвистическое значение.

Естественным шагом при построении функций принадлежности элементов терм-множества лингвистической переменной является построение одновременно всех функций принадлежности этого терм-множества, сгруппированных в так называемое отношение моделирования R [1]. Процесс построения состоит в заполнении таблицы, где, например, для лингвистической переменной «расстояние» столбцы индексированы расстояниями в метрах, а строки — элементами терм-множества «очень близко», «близко», ..., «далеко», «очень далеко». На пересечении соответствующей строки и столбца стоит степень сходства для испытуемого эксперта данных понятий между собой в данной семантической ситуации, например, насколько сходны понятия «5 метров» и «близко» в ситуации перебегания улицы перед быстро идущим транспортом. Расстояние берётся от пешехода до машины и в данном случае является синонимом опасности. Вообще говоря, каждую клетку таблицы можно заполнять отдельно, а потом, переставляя строки и столбцы, постараться сделать функции принадлежности *унимодальными*. Если это удается, то исходное терм-множество может быть использовано для построения нечёткой шкалы измерений, точками отсчёта которой являются сами элементы терм-множества. Перевод в эту шкалу осуществляется с помощью минимаксного умножения строки, задающей исходную лингвистическую переменную в шкале метров, на отношение моделирования. Отношение сходства между элементами терм-множества $R * R^T$, полученное с помощью умножения матрицы R на транспонированную саму себя, задаёт набор функций принадлежности элементов лингвистической шкалы в самой шкале, а отношение $R^T * R$ задаёт набор функций принадлежности расстояний в метрах в метрической шкале.

1.1.6. Использование источников знаний третьего рода для извлечения нечёткости

В ряде работ [23, 8] приводится метод построения функций принадлежности на основе собранных статистических данных, которые могут в свою очередь находиться в базах данных (*источниках знания третьего рода*). В приводимом далее методе в качестве степени принадлежности элемента какому-либо нечёткому множеству принимается оценка частоты использования понятия, которое описывается соответствующим нечётким множеством. В рассматриваемом методе получаются достаточно гладкие функции принадлежности, так как используются специально разработанные механизмы — матрицы подсказок.

Рассматриваемый метод целесообразно использовать в тех областях, где накоплен человеческий опыт, выраженный в статистических данных о каких-то событиях. Например, при построении автоматизированных систем управления возникает задача моделирования деятельности человека-оператора. С этой задачей справляется аппарат нечёткой логики, в частности рассматриваемый метод позволяет в автоматизированном режиме получать описания функций принадлежности.

В своей деятельности относительно решения задач человек не пользуется конкретными числами для оценки тех или иных явлений. Вместо этого он использует значения лингвистических переменных. Каждое значение лингвистической переменной (то есть нечёткая переменная — см. Глоссарий) описывается определённой функцией принадлежности, которая индивидуальна для каждого человека.

Пусть в некотором эксперименте человек-оператор n раз фиксирует своё внимание на том, имеет ли место факт A или нет. Событие, заключающееся в n проверках наличия факта A, называется оценочным. Пусть в k проверках факт A имел место, в этом случае человек оператор фиксирует частоту $p = k / n$ появления факта A и оценивает её с помощью слов «часто», «редко» и т. п. В этом эксперименте имеет место лингвистическая переменная «Частота проявления факта A» (или, упрощая, просто «Частота»), значениями которой являются нечёткие лингвистические термы «часто», «редко» и др.

Оценивая частоту p, человек-оператор опирается на свой опыт, который отражает частоту появления факта A в событиях прошлого, представляющихся оператору аналогичными оцениваемому событию. К нему также поступает информация, основанная на наблюдении других людей, а также информация, отражающая общественный опыт вообще. В зависимости от степени доверия к источнику такого рода информации в базах данных она запоминается с различными весами.

Степень принадлежности некоторого значения конкретной функции принадлежности вычисляется как отношение числа экспериментов, в которых исследуемое значение встретилось в определённом интервале шкалы (соответствующем функции принадлежности), к максимальному для этого значения числу экспериментов по всем интервалам. Метод основывается на условии, что в каждый интервал шкалы попадает одинаковое число экспериментов. Конечно, в реальных условиях это предположение соблюдается очень редко, поэтому составляется эмпирическая таблица, в которой эксперименты могут быть неравномерно распределены по интервалам.

Пусть на основе экспериментов в базе данных хранятся следующие значения частоты появления лингвистических термов:

Таблица 3. Оценка отклонения параметра технологического процесса в терминах лингвистической переменной «Относительная величина»

Значение	Интервал																			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
ОЧЕНЬ МАЛО	3	7	3	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
МАЛО	0	0	1	0	4	1	6	4	1	0	0	0	0	0	0	0	0	0	0	0
СРЕДНЕ	0	0	0	0	0	0	0	2	2	5	7	0	0	0	0	0	0	0	0	0
МНОГО	0	0	0	0	0	0	0	0	0	0	3	8	0	7	5	2	3	0	0	0
ОЧЕНЬ МНОГО	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	5	7	5	2

Используя свойства функций принадлежности, необходимо предварительно обработать данные из представленной таблицы таким образом, чтобы уменьшить искажения, вносимые экспериментом. Естественными свойствами функций принадлежности являются наличие одного максимума и гладкие, затухающие до нуля фронты [12, 8, 20]. Для обработки статистических данных можно воспользоваться так называемой *матрицей подсказок*. Предварительно из статистических данных удаляются явно ошибочные элементы (например, элемент ОЧЕНЬ МАЛО в интервале 17 в представленной таблице). Критерием удаления служит наличие нескольких нулей в строке вокруг удаляемого элемента.

Элементы матрицы подсказок вычисляются по формуле:

$$k_j = \sum_{i=1}^{n} b_{ij} \ ,$$

где n — число нечётких лингвистических термов (в рассматриваемом примере — 5). При этом параметр j изменяется от 1 до количества интервалов (в рассматриваемом примере — 20). Таким образом, матрица подсказок для представленной таблицы выглядит следующим образом:

$$K = \begin{Vmatrix} 3 & 7 & 4 & 0 & 5 & 1 & 6 & 6 & 3 & 5 & 10 & 8 & 0 & 7 & 6 & 4 & 9 & 7 & 5 & 2 \end{Vmatrix}.$$

В каждой строке таблицы выбирается максимальный элемент $k_{max} = \max k_j$, после чего все элементы исходной таблицы преобразуются по формуле:

$$c_{ij} = \frac{b_{ij} k_{\max}}{k_j}, \ i = \overline{1,5}, \ \ j = \overline{1,20} \, .$$

Для столбцов, где $k_j = 0$, применяется линейная аппроксимация:

$$c_{ij} = \frac{c_{ij-1} + c_{ij+1}}{2}, \ i = \overline{1,5}, \ \ j = \overline{1,20} \, .$$

Для построения функции принадлежности в исходной таблице находятся максимальные элементы по строкам, по чьему индексу определяется фактический максимальный элемент новой таблицы. Значение функции принадлежности вычисляется по формуле:

$$\mu_{ij} = \frac{c_{ij}}{c_{i\max}},\ c_{i\max} = \max_j c_{ij},\ i = \overline{1,5},\ \ j = \overline{1,20}.$$

На следующем рисунке показаны рассчитанные функции принадлежности для нечётких переменных из представленной ранее таблицы:

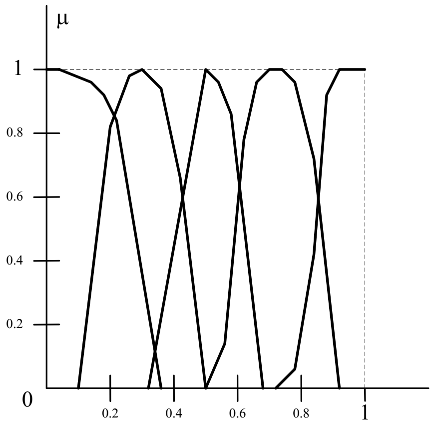

Рисунок 6. Функции принадлежности значений лингвистической переменной «Относительная величина»

Слева направо показаны графики функций принадлежности нечётких переменных «Очень мало», «Мало», «Средне», «Много» и «Очень много». Видно, что все функции принадлежности удовлетворяют описанным выше критериям.

1.1.7. Параметрический подход к построению функций принадлежности

В случае если существует необходимость в построении модифицированных функций принадлежности на основе имеющихся, то можно воспользоваться параметрическим подходом [20]. По сути, этот подход соответствует компоненту M модели лингвистической переменной (см. Глоссарий), то есть осуществляет построение функций принадлежности неизвестных нечётких термов на основе известных.

Рассматриваемый метод работает только с функциями принадлежности, построенными по трём точкам, то есть на шкале, представляющей область определения нечёткой переменной (или на универсальной шкале), отмечаются три точки:

1. Точка, которая ещё не принадлежит рассматриваемой нечёткой переменной (точка A);
2. Точка, которая типична для рассматриваемой нечёткой переменной (точка B);
3. Точка, которая уже не принадлежит рассматриваемой нечёткой переменной (точка C).

В этом случае функция принадлежности имеет следующий вид:

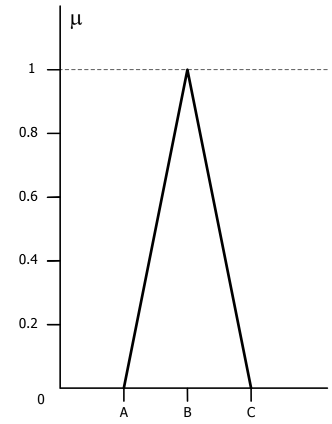

Рисунок 7. Типичный вид функции принадлежности

Такое представление функций принадлежности лингвистических термов называется параметрическим. Одним из видов такого представления является S-образная функция, когда одним концом она уходит в бесконечность того или иного знака, то есть либо точка A лежит в отрицательной бесконечности, либо точка C лежит в бесконечности положительной. Таким образом, выделяется три типичных формы функций принадлежности лингвистических термов [12, 8].

Для того чтобы получить начальные данные для рассматриваемого метода, необходимо, чтобы для лингвистической переменной рассмотренным выше образом было построено две функции принадлежности, причём одна из функций принадлежности должна представлять собой модификацию другой. То есть если есть два нечётких лингвистических терма t и t’, то они должны быть связаны отношением: t’ = h(t). При этом h — ограничение (модификатор) на t типа ДОВОЛЬНО, БОЛЕЕ-МЕНЕЕ, НЕ ОЧЕНЬ и т. п.

Задача параметрического построения функций принадлежности состоит в описании и последующем использовании функции перехода от t к t’. Функция перехода находится при помощи параметров термов t (z_1, z_2, z_3) и t’ (w_1, w_2, w_3), при этом считается, что параметры упорядочены отношением «меньше».

В случае S-образных функций принадлежности задача решается тем же способом, просто полагается, что для соответствующей функции либо первый, либо последний параметр стремится бесконечности соответствующего знака.

Для того чтобы построить функцию перехода, необходимо воспользоваться аппаратом автоморфных функций. Рассматривается дробно-линейное преобразование прямой в себя, вида:

$$T : x \to \frac{\alpha x + \beta}{\gamma x + \delta}.$$

Это преобразование удобно расширить, включив в ось действительных чисел **R** точку ∞. При этом необходимо условиться, что T (-δ/γ) = ∞ и T (∞) = α/γ. В этом случае оказывается, что рассмотренное дробно-линейное преобразование взаимно однозначно отображает расширенную прямую $\mathbf{R} \cup \{\infty\}$ на себя. Бесконечное множество таких преобразований, где α, β, γ и δ суть действительные числа, представляет собой так называемую модулярную группу: обратные преобразования и произведения дробно-линейных отображений также являются дробно-линейными.

Преобразование T^{-1}, обратное к T, получается, если уравнение

$$z = \frac{\alpha w + \beta}{\gamma w + \delta}$$

разрешить относительно w. Таким образом, обратное преобразование T^{-1} имеет вид:

$$T^{-1} : w = \frac{-\delta z + \beta}{\gamma z - \alpha}.$$

При параметрическом представлении функций принадлежности нечётких лингвистических термов задача описания перехода от одного терма t (z_1, z_2, z_3) к другому t’ (w_1, w_2, w_3) сводится к непосредственному подсчёту коэффициентов дробно-линейного преобразования по формулам:

$$\begin{gathered}
\alpha = z_1 z_2 (w_1 - w_2) + z_1 z_3 (w_3 - w_1) + z_2 z_3 (w_2 - w_3) \\
\beta = w_1 w_2 z_3 (z_1 - z_2) + w_1 w_3 z_2 (z_3 - z_1) + w_2 w_3 z_1 (z_2 - z_3) \\
\gamma = z_2 (w_1 - w_3) + z_1 (w_3 - w_2) + z_3 (w_2 - w_1) \\
\delta = w_1 w_2 (z_1 - z_2) + w_1 w_3 (z_3 - z_1) + w_2 w_3 (z_2 - z_3)
\end{gathered}.$$

Эти же коэффициенты определяют и обратный переход от t’ к t.

В случае если необходимо осуществить переход от S-образной ($z_1 = \infty$) к треугольной функции принадлежности, то следует воспользоваться упрощёнными формулами расчёта коэффициентов:

$$\begin{gathered}
\alpha = z_2 (w_2 - w_1) + z_3 (w_1 - w_3) \\
\beta = w_2 z_3 (w_3 - w_1) + w_3 z_2 (w_1 - w_{2)} \\
\gamma = w_2 - w_3 \\
\delta = w_1 (w_3 - w_2)
\end{gathered}.$$

Если же у S-образной функции принадлежности третий параметр равен бесконечности ($z_3 = \infty$), то в этом случае формулы для расчёта коэффициентов функции перехода также видоизменяются:

$$\begin{gathered}
\alpha = z_1 (w_1 - w_3) + z_2 (w_3 - w_2) \\
\beta = w_1 z_2 (w_2 - w_3) + w_2 z_1 (w_3 - w_1) \\
\gamma = w_1 - w_2 \\
\delta = w_3 (w_2 - w_1)
\end{gathered}.$$

Остаётся последний случай — переход от одной S-образной функции, к другой (по сути, оба нечётких терма описываются одной наклонной прямой, вернее, её отрезком). В этом случае имеет место обычное линейное преобразование:

$$L : x \to \alpha x + \beta.$$

Коэффициенты перехода в этом случае будут иметь следующий вид (необходимо отметить, что это уже не параметры функций принадлежности, а точки пересечения прямой, определяющей нечёткий терм с ординатами 0 и 1):

$$\alpha = \frac{y_2 - y_1}{x_2 - x_1}$$
$$\beta = \frac{y_1 x_2 - y_2 x_1}{x_2 - x_1}.$$

В качестве примера использования рассмотренного метода можно привести проблему автоматического определения функции принадлежности нечётких термов в следующей задаче:

Пусть экспертом составлены параметрические описания нечётких термов «ПРОХЛАДНАЯ» и «НЕ ПРОХЛАДНАЯ» для лингвистической переменной «Температура воды для купания». Необходимо построить функцию перехода для модификатора «НЕ» и применить её к нечёткому терму «ЛЕДЯНАЯ», чтобы автоматически получить описание нечёткого терма «НЕ ЛЕДЯНАЯ».

Пусть эксперт задал следующие параметры функций принадлежности для рассматриваемых термов:

- ПРОХЛАДНАЯ: ($z_1 = \infty$, $z_2 = 14$, $z_3 = 16$)
- НЕ ПРОХЛАДНАЯ: ($w_1 = 17$, $w_2 = 18$, $w_3 = 20$)
- ЛЕДЯНАЯ: ($z_1 = \infty$, $z_2 = 4$, $z_3 = 6$)

По нечётким термам «ПРОХЛАДНАЯ» и «НЕ ПРОХЛАДНАЯ» находятся коэффициенты функции перехода:

$\alpha = z_2(w_2 - w_1) + z_3(w_1 - w_2) = 14 * (18 - 17) + 16 * (17 - 20) = -34$
$\beta = w_2 z_3(w_3 - w_1) + w_3 z_2(w_1 - w_2) = 18 * 16 * (20 - 17) + 20 * 14 * (17 - 18) = 584$
$\gamma = w_2 - w_3 = 18 - 20 = -2$
$\delta = w_1(w_3 - w_2) = 17 * (20 - 18) = 34$

По найденным коэффициентам теперь легко можно построить параметрическую функцию принадлежности для нечёткого терма «НЕ ЛЕДЯНАЯ»:

$z_1 = \infty$; $w_1 = 34 / 2 = 17$;
$z_2 = 4$; $w_2 = (-34 * 4 + 584) / (-2 * 4 + 34) = 17.2$;
$z_3 = 6$; $w_3 = (-34 * 6 + 584) / (-2 * 6 + 34) = 17.3$.

Таким образом, параметрическое представление функции принадлежности нового нечёткого терма «НЕ ЛЕДЯНОЙ» выглядит так: ($w_1 = 17$, $w_2 = 17.2$, $w_3 = 17.3$).

1.2. Извлечение знаний с элементами неопределённости

При решении реальных задач эксперты редко бывают уверены на 100 % относительно каких-либо своих утверждений [48], таким образом, при разработке систем, основанных на знаниях эту неуверенность эксперта каким-либо образом необходимо формализовать. Обычно такая неопределённость описывается при помощи числа из интервала [0; 1], и это число называется «субъективная вероятность», «фактор уверенности» и др.

Главная проблема в таком подходе заключается в том, что необходимо использовать точное число для описания неопределённости. Эксперт может довольно легко провести различия между описаниями неопределённости 0.9 и 0.5, однако для эксперта будет весьма сложно сделать заключение относительно разницы факторов уверенности 0.7 и 0.701. При извлечении неопределённых знаний эксперту для описания своей уверенности относительно тех или иных фактов проще оперировать не точными числами из интервала [0; 1], а некоторыми непрерывными подинтервалами из этого интервала. В работе [48] утверждается, что на таком интервале неопределённости можно даже построить функцию $\mu_d(A)$, которая каждому числу d из своей области определения ставит в соответствие степень уверенности в том, что это число является степенью уверенности эксперта в истинности факта A. Таким образом, можно сказать, что функция $\mu_d(A)$ описывает неопределённость второго порядка.

Теоретически таким же образом можно определить неопределённость третьего, четвёртого и т. д. порядка, однако использование на практике таких описаний чрезвычайно затруднительно, так как их сложно обрабатывать, а к тому же непросто найти приложения, где такие описания неопределённости можно было бы адекватно использовать. Тем более что извлекать из эксперта знания с неопределённостью порядка выше второго сложнее, чем знания с неопределённостью первого порядка. На практике используется неопределённость порядка не выше второго [46].

Таким образом, для извлечения знаний с неопределённостью у эксперта достаточно спрашивать его уверенность относительно используемых им в процессе рассуждений фактов, понятий и отношений в виде подинтервалов интервала [0; 1]. При этом у эксперта обычно не возникает никаких сложностей с таким неточным описанием своей неопределённости [12].

В работе [10] приводится новый формализм, на основе которого можно извлекать неопределённые знания из эксперта. Этот формализм называется *алгебраической байесовской сетью* и является редуцированным вариантом вероятностного пространства. Уже само название показывает, что рассматриваемый подход основан на теории вероятности и статистических методах. Отмечается, что такие методы являются наиболее строгими и глубоко проработанными. Благодаря закону больших чисел они имеют твёрдые позиции в практике моделирования неопределённых знаний.

Однако классические вероятностные модели не приспособлены к описанию неопределённых знаний эксперта. В работах [39, 52] обобщаются вероятностные модели таким образом, что стало возможным оперировать интервальными оценками неопределённости, которые задаются нижней и верхней границей. Фактически это означало появление новых моделей неопределённости знаний, которые не требуют от экспертов выражать свои знания утверждениями с точными оценками мер истинности.

В качестве примера можно привести экспертные утверждения о зависимостях в проблемной области. Пусть при извлечении знаний экспертам предлагается высказать свои суждения о зависимости между парой утверждений. Известно, что такие суждения обычно носят качественный и неформальный характер. Имеется довольно ограниченный набор высказываний экспертов для описания зависимостей: «x_1 характерно для x_2», «x_1 наблюдается при x_2», «x_1 часто сопровождает x_2», «x_1 может наблюдаться при x_2» и т. п. Очевидно, что некоторые из таких высказываний близки и трудно различимы по смыслу, и поставить в соответствие каждому высказыванию конкретную формулу вероятностной логики невозможно. Но опрос экспертов можно построить так, что итогом будет подмножество формул с интервальной мерой истинности, например:

$$\begin{array}{l} p(x_1) \in [0.3,\ 0.9] \\ p(\overline{x_1}) \in [0.1,\ 0.8] \\ p(\overline{x_1} \to x_2) \in [0.2,\ 0.9] \\ p(x_1 \to \overline{x_2}) \in [0.2,\ 0.7] \end{array},$$

к которому могут быть добавлены некоторые качественные ограничения:

$$\begin{array}{l} p(x_1) \geq p(x_2) \\ p(\overline{x_1}) \geq p(x_1) \\ p(x_1 \,\&\, \overline{x_2}) \geq p(x_1 \,\&\, x_2) \\ p(\overline{x_1}) \geq 3p(x_1 \,\&\, x_2) \end{array}.$$

Это — характерный пример того, когда следует использовать модель интервальных вероятностей истинности формул.

В работе [30] предлагается лингвистический подход к извлечению знаний с неопределённостью, суть которого состоит в выявлении в естественно-языковых высказываниях эксперта некоторых ключевых слов, которые могут сигнализировать о присутствии в высказывании неопределённости.

В естественном языке существуют так называемые размытые квантификаторы [23], поэтому наличию во фразах информации о неопределённости соответствует использование в речи достаточно ограниченного количества специальных слов и словосочетаний. Каждому такому ключевому слову (или ключевой фразе) соответствует определённая служебная информация, записанная в словаре лингвистического процессора, производящего обработку входных естественно-языковых фраз, при помощи которой последний строит формализованное описание неопределённости. В соответствии с тем, как впоследствии будет использоваться отловленная информация, каждому ключевому слову может быть поставлено в соответствие либо конкретное число из интервала [0; 1], либо конкретное непрерывное подмножество того же интервала.

1.2.1. Применение метода репертуарных решёток для извлечения неопределённости

В середине XX века появилась работа Дж. Келли «Психология личных конструктов», где излагались оригинальная теория личности и новый метод исследования личности. Дж. Келли создал теорию личности, которая позже была названа «теория личных (или личностных) конструктов», а также разработал методику её практического применения — метод «репертуарных решёток». Теория и метод, предложенные Дж. Келли для врачебной психотерапевтической практики, получили новый импульс для их практического использования в искусственном интеллекте, и, в первую очередь, при создании интеллектуальных систем, основанных на знании [24]. Краткая характеристика этому своеобразному, интересному и многообещающему методу может быть взята, например, из работы [32].

В основе теории конструктов лежит идея о том, что каждый человек представляет собой исследователя. Дж. Келли полагал, что любой человек в течение всей жизни ищет смысл в себе и в окружающей обстановке. Для этого он создаёт и перестраивает свои собственные системы взглядов, выдвигает гипотезы, проверяет их на практике, корректирует, вносит изменения в теорию (изменяет своё мышление) и так до бесконечности. Дж. Келли разработал технику *репертуарных решёток* в качестве метода изучения систем личных конструктов другого человека, как способ «влезть в шкуру другого человека», увидеть мир его глазами, а также «увидеть то, что стоит за словами», то есть выявить те самые глубинные семантические связи внутри проблемной области, которые эксперту практически невозможно вербализовать.

Репертуарная решётка представляет собой матрицу, которая заполняется либо самим испытуемым экспертом, либо инженером по знаниям в процессе беседы. Столбцам матрицы соответствует группа элементов, объектов — значимых понятий из конкретной проблемной области. Строки матрицы представляют собой конструкты — биполярные признаки, параметры или шкалы. Конструкты либо задаются инженером по знаниям, либо выявляются с помощью специальных процедур.

Понятие конструкта — центральное в теории Дж. Келли. Было замечено, что когда человек говорит: «что-то является чем-то», он всегда неявно предполагает, чем это что-то одновременно не является. Именно конкретное расположение оценки данного элемента на оси «является чем-то — не является чем-то» определяет индивидуальное восприятие этого элемента конкретным человеком. Совокупность конструктов представляет собой набор значимых осей, относительно которых человек рассматривает и оценивает свой и окружающий мир. В процессе заполнения репертуарной решётки испытуемый оценивает каждый элемент по каждому конструкту. Далее заполненная решётка подвергается статистической обработке.

Анализ репертуарной решётки позволяет оценить силу и направленность связей между конструктами, выявить наиболее значимые конструкты (глубинные), а также иерархические отношения между конструктами. В следующей таблице приведена заполненная оценочная решётка (для оценки используется 11-балльная шкала), позволяющая сравнить между собой восемь различных теорий исследований личности по восьми параметрам, значимых для обоснования теорий. Элементы этой решётки — фамилии учёных, авторов теорий, конструкты — теоретические положения (пример взят из работы [32]).

Таблица 4. Пример репертуарной решётки для сравнения восьми теорий исследования личности

	Фрейд	Эриксон	Мюррей	Скиннер	Олпорт	Келли	Маслоу	Роджерс	
Свобода	11	8	10	11	5	6	2	1	Детерминизм
Рациональность	10	3	2	6	1	1	2	1	Иррациональность
Холизм	3	1	3	11	3	3	1	1	Элементаризм
Наследственность	3	10	6	11	6	9	5	4	Окружающая среда
Субъективность	5	8	4	11	5	1	2	1	Объективность
Активность	4	3	4	11	1	6	1	1	Реактивность
Гомеостаз	1	9	2	6	10	6	10	11	Гетеростаз

Познаваемость	1	4	6	1	4	11	11	11	Непознаваемость

Обработка этой решётки с помощью статистических методов и анализ интеркорреляций конструктов позволяет прийти к выводам типа: теории, делающие акцент на свободе человека в противоположность детерминизму, не имеют в своей основе допущений о «познаваемости» и «гомеостазе» (интеркорреляция соответствующих элементов отрицательна), однако они базируются на представлениях о «субъективности», «рационализме» и «активности» (высокие интеркорреляции).

Таблица 5. Интеркорреляция конструктов при сравнении методов исследования личности

Свобода/Детерминизм	1	0.69	0.58	0.34	0.69	0.68	-0.82	-0.82
Рациональность/Иррациональность	0.69	1	0.40	-0.13	0.49	0.42	-0.64	-0.73
Холизм/Элементаризм	0.58	0.40	1	0.54	0.70	0.90	-0.30	-0.56
Наследственность/Окружающая среда	0.34	-0.13	0.54	1	0.61	0.67	0.10	-0.21
Субъективность/Объективность	0.69	0.49	0.70	0.61	1	0.62	-0.18	-0.85
Активность/Реактивность	0.68	0.42	0.90	0.67	0.62	1	-0.46	-0.45
Гомеостаз/Гетеростаз	-0.82	-0.64	-0.30	0.10	-0.18	-0.46	1	-0.48
Познаваемость/Непознаваемость	-0.82	-0.73	-0.56	-0.21	-0.85	-0.45	-0.48	1

Метод репертуарных решёток сейчас — это целое направление в практической психодиагностике. Имеются исследования артикулированности (степени структурированности и связности) систем конструктов, показавшие, например, качественные различия в решётках, заполненных здоровыми людьми и больными неврозами. Интенсивность — другой показатель, позволяющий по данным решётки дифференцировать людей по степени «рыхлости» системы конструктов. «Рыхлая» система конструктов свидетельствует о том, что человек в данной области не способен чётко мыслить и планировать свои действия.

Также есть и другие показатели репертуарной решётки, имеющие к теме извлечения знаний непосредственное отношение (например, когнитивная сложность — способность оценивать внешний мир одновременно по определённому количеству параметров и др.).

После статистической обработки системы конструктов можно выделить коэффициенты интеркорреляции, которые в свою очередь можно трактовать как коэффициенты уверенности в том, что один конструкт проявляется вместе с другим. То есть таким образом можно автоматически получать описание неопределённости. Например, по данным из предыдущей таблицы можно заключить, что значение на шкале «Свобода / Детерминизм» проявляется вместе со значением со шкалы «Субъективность / Объективность» с уверенностью в 0.69.

Таким образом, теория личных конструктов является тем методом, который ориентирован на решение задачи извлечения знаний из эксперта с учётом его собственного, специфического, субъективного видения проблемной ситуации. Одновременно метод позволяет оценить уровень соответствия испытуемого тем требованиям, которые предъявляются к «идеальному эксперту». По всей видимости, решётку можно рассматривать как разновидность структурированного интервью, во время которого неявно извлекаются глубинные взаимосвязи между понятиями, которыми оперирует эксперт в процессе решения задачи. При этом с помощью статистических методов обработки можно получать коэффициенты неопределённости, относящиеся к взаимосвязям понятий в исследуемой проблемной области.

1.3. Извлечение неточных и недоопределённых знаний

С точки зрения извлечения знаний неточность и недоопределённость в какой-то мере простые НЕ-факторы. Оба этих НЕ-фактора могут проявляться в рассуждениях эксперта в виде интервалов, внутри которых содержится извлекаемое значение некоторой конкретной переменной (параметра, атрибута).

В отношении неточности можно сказать, что неточные значения проявляются тогда, когда в процессе решения задачи используются значения, полученные при помощи измерительных приборов с оговорённой погрешностью измерения. Такая погрешность обуславливает наличие интервала, с точностью до которого можно указать значение измеренного параметра.

Таким образом, неточность можно извлекать напрямую из эксперта, спрашивая его об относительной или абсолютной погрешности прибора, если описываемый в процессе извлечения знаний параметр был получен при помощи измерительного прибора. Обычно у эксперта не возникает сложностей с описанием погрешности [12]. Кроме того, эксперт может сразу задать интервал, в котором находится значение измеренного параметра, а также вид функции распределения вероятности нахождения этого точного значения внутри интервала. Такую функцию распределения вероятности можно в дальнейшем использовать для преобразования неточности в нечёткость (см. Глава 3).

Недоопределённость проявляется при частичном отсутствие знаний о значении какого-либо параметра (не важно, измеримого или нет) [21]. В случае измеримых параметров недоопределённость и неточность можно легко приводить друг к другу, однако существует довольно важное различие. В случае недоопределённости частичное отсутствие знаний можно восполнять, постепенно доопределяя параметр, а неточные измеренные параметры самодостаточны сами по себе, так как зачастую повышать точность измерения для решения конкретной задачи не имеет смысла.

То есть можно утверждать, что недоопределённые знания используются в алгоритмах решения задач, когда ограничения на значение искомого параметра постепенно суживают область поиска достоверного значения. В случае если искомый параметр является числовой величиной, то ограничения на его значения выражаются при помощи интервалов. Как уже было отмечено, обычно у экспертов не возникает проблем при извлечении интервальных величин.

Другой подход к извлечению неточности и недоопределённости заключается в использовании лингвистических методов [30]. Иногда информацию о неточности и недоопределённости можно получить непосредственно из естественно-языковых высказываний эксперта. Хотя этот случай довольно сложен с точки зрения обработки, он может быть реализован при помощи техники поиска ключевых слов.

Во вводимых экспертом естественно-языковых фразах производится поиск определённых слов, которые могут свидетельствовать о наличии информации о рассматриваемых НЕ-факторах во введённых фразах. Если такие слова обнаруживаются, то производится создание предварительных описаний, которые помещаются в поле знаний, которое затем исправляется на этапе верификации инженером по знаниям. Этот метод может быть хорошим подспорьем в автоматизированном режиме приобретения знаний, так как позволяет незаметно для эксперта (соответственно не утруждая его дополнительным вопросами) формировать структуры, описывающие НЕ-факторы с большой степенью уверенности, которые впоследствии обрабатывает инженер по знаниям.

Основная и главная проблема в отношении извлечения и приобретения неточных и недоопределённых знаний заключается в том, что на текущий момент эти два НЕ-фактора формализованы очень слабо. Если недоопределённость можно в каком-то приближении формализовать при помощи ограничений и теории программирования в ограничениях (*constraint programming*) [21], то относительно неточности сложно сделать какие-либо заключения. К тому же многие исследователи подразумевают под неточностью совершенно различные понятия [22, 34].

Глава 2. Представление знаний с НЕ-факторами

В соответствии с определением процесса представления знаний, сформулированным в Глоссарии, при представлении знаний с НЕ-факторами необходимо ответить на два вопроса: «Что представлять?» и «Как представлять?».

Ответ на первый вопрос довольно простой: представлять надо информацию о НЕ-факторах, полученную в ходе приобретения знаний. С другой стороны не всегда ясно, какая именно информация о НЕ-факторах пришла с выхода процесса приобретения знаний, однако эта проблема должна решаться до начала построения системы, основанной на знаниях, поэтому в дальнейшем будет считаться, что формализмы, описывающие те или иные НЕ-факторы, известны как на этапе приобретения знаний, так и на дальнейших этапах.

Вопрос о «физическом» представлении приобретённой информации о НЕ-факторах необходимо рассматривать с нескольких точек зрения:

- возможность быстрого исправления введённой информации, что подразумевает использование формата данных, который может быть воспринят человеком;
- структурированность информации, её адекватность и интуитивное понимание не должны выходить за рамки здравого смысла;
- эффективная обработка представленной информации о НЕ-факторах, то есть представление должно быть эффективным с точки зрения адекватности описания НЕ-фактора, но в то же время простым с точки зрения компьютерной обработки.

В дальнейшем изложении для каждого выбранного НЕ-фактора будут приведены ответы на два поставленных вопроса о представлении знаний, а также приведены и по возможности сравнены различные математические аппараты собственно для представления знаний с НЕ-факторами.

2.1. Представление нечёткости

Нечёткость представляет собой наиболее изученный НЕ-фактор, для которого с самого начала был предложен математический формализм. Таким формализмом является функция принадлежности нечёткого множества, которая соответствует характеристической функции обычного множества [23]. Классическое определение функции принадлежности нечёткого множества A выглядит следующим образом [57]:

$$\mu_A(x) : U \to [0;1],$$

где U — универсум (универсальное множество). Следовательно, относительно некоторых элементов универсального множества невозможно однозначно сказать принадлежат эти элементы нечёткому множеству A или нет, как это можно сделать в классической теории множеств.

Таким образом, ответ на вопрос «Что представлять?» в случае нечёткости однозначен — представлять необходимо функции принадлежности нечётких множеств. Однако ответ на вопрос «Как представлять?» не так однозначен, и до сих пор не существует формализма, который сочетал бы в себе простоту и эффективность работы для любой проблемной области, для которой создаётся база знаний с элементами нечёткости. Но существуют методы представления, для которых отношение простота/эффективность практически достигает оптимального значения. Такими методами являются нечёткие числа LR-типа [18] для представления нечётких множеств, определённых на действительной оси **R**, а также кусочно-линейные функции принадлежности для любых нечётких множеств.

2.1.1. Использование нечётких чисел LR-типа

Нечёткие числа — в общем случае это нечёткие переменные, определённые на числовой оси действительных чисел **R** [18]. Нечёткое число можно рассматривать как нечёткое множество A на множестве действительных чисел:

$$\mu_A(x) \in [0; \quad 1], \ x \in R.$$

Нечёткие числа LR-типа — это разновидность нечётких чисел специального вида, то есть описываемых определёнными правилами с целью снижения объёма вычислений при операциях над такими числами.

Функции принадлежности нечётких чисел LR-типа задаются при помощи двух функций действительного переменного L (x) и R (x), которые удовлетворяют следующим свойствам:

- L(-x) = L(x)
- R(-x) = R(x)
- L(0) = R(0)

При этом отмечается, что максимум обеих функций равен 1 и достигается в точке 0.

В свою очередь нечёткие числа LR-типа делятся на унимодальные и толерантные. Унимодальное нечёткое число имеет одну и только одну точку, где функция принадлежности этого нечёткого числа принимает значение 1. Функция принадлежности толерантного нечёткого числа принимает значение 1 на некотором интервале, состоящим более чем из одной точки.

Пусть имеются две функции L (x) и R (x), которые удовлетворяют поставленным требованиям. Тогда унимодальное нечёткое число будет определяться тремя параметрами:

$$\mu_A(x) = \begin{cases} L\left(\dfrac{a-x}{\alpha}\right), & если \quad x \le a \\ R\left(\dfrac{x-a}{\beta}\right), & если \quad x > a \end{cases},$$

где a — мода; $\alpha \geq 0$, $\beta \geq 0$ — левый и правый коэффициенты нечёткости соответственно (эти коэффициенты могут трактоваться как «пологость» соответствующей функции).

Для толерантных нечётких чисел необходимо четыре параметра: a_1, a_2, α и β, где a_1 и a_2 — границы толерантности, то есть в интервале $[a_1; a_2]$ функция принадлежности нечёткого числа принимает значение 1.

В работах [12, 23, 8] отмечается, что решение задач математического моделирования сложных систем с применением аппарата нечётких множеств требует выполнения большого объёма операций над лингвистическими переменными. Для удобства исполнения операций, а также для ввода/вывода и хранения данных, желательно работать с функциями принадлежности нескольких стандартных типов.

Нечёткие множества, которыми приходится оперировать в большинстве задач, являются, как правило, унимодальными и нормальными. Одним из возможных методов аппроксимации унимодальных нечётких множеств является аппроксимация с помощью функций принадлежности LR-типа.

2.1.1.1. Триангулярные и трапециевидные числа LR-типа

Специальным видом нечётких чисел LR-типа являются триангулярные (треугольные) и трапециевидные нечёткие числа, которые в основном используются при представлении лингвистических переменных в задачах управления, проектирования и планирования [40]. Триангулярные нечёткие числа используются в тех случаях, где необходимы унимодальные числа. В свою очередь трапециевидные нечёткие числа являются толерантными.

Аппарат триангулярных и трапециевидных нечётких чисел LR-типа был разработан для оптимизации количества вычислений, связанных с обработкой нечёткости при решении различных задач, так как эти виды нечётких чисел имеют чрезвычайно простое представление, что резко уменьшает количество и сложность вычислений, связанных с их обработкой.

Триангулярное нечёткое число может быть представлено в виде тройки:

$$TFN = \langle a, b, c \rangle,$$

где:

- a — та точка на действительной оси **R**, которая ещё не принадлежит триангулярному нечёткому числу;
- b — точка, где функция принадлежности нечёткого числа достигает максимума (мода);
- c — точка, которая уже не принадлежит триангулярному нечёткому числу (см. следующую диаграмму).

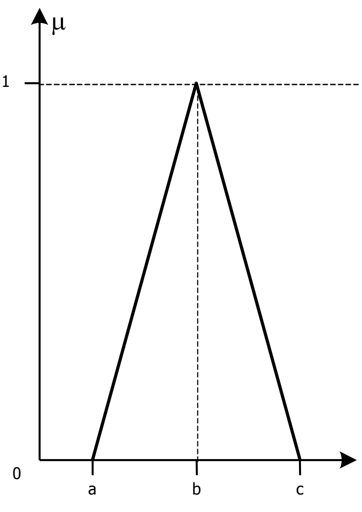

Рисунок 8. Триангулярное нечёткое число LR-типа

В свою очередь трапециевидное нечёткое число может быть представлено в виде четвёрки:

$$TrFN = \langle a, b, c, d \rangle,$$

где:

- a — та точка на действительной оси **R**, которая ещё не принадлежит трапециевидному нечёткому числу;
- b — точка, где функция принадлежности нечёткого числа достигает максимума (начало области толерантности);
- c — точка, где заканчивается область максимума нечёткого числа;
- d — точка, которая уже не принадлежит трапециевидному нечёткому числу (см. следующую диаграмму);

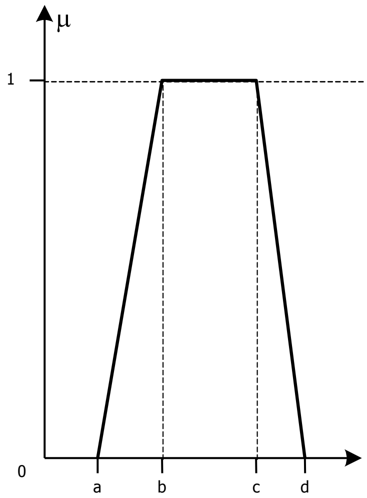

Рисунок 9. Трапециевидное нечёткое число LR-типа

Интервал [b; c] на области определения трапециевидного нечёткого числа называется зоной толерантности.

2.1.1.2. Шеститочечное представление нечётких чисел

Шеститочечное представление нечёткого числа основано на шести точках на действительной оси **R**. То есть представление нечёткого числа в таком виде выглядит следующим образом [42]:

$$FN6 = \left\langle \underline{m}^{\varepsilon}, \underline{m}^{\lambda}, \underline{m}, \overline{m}, \overline{m}^{\lambda}, \overline{m}^{\varepsilon} \right\rangle,$$

где числа ε и λ рассматриваются с точки зрения построения функции принадлежности такого нечёткого числа экспертом:

- $\alpha = 1$, $\mu(x) = 1$ — значение x полностью принадлежит рассматриваемому нечёткому множеству.
- $\alpha = \lambda$, $\mu(x) > \lambda$ — эксперт ожидает, что значение x с $\mu(x) \geq \lambda$ имеет неплохой шанс принадлежать рассматриваемому нечёткому множеству.
- $\alpha = \varepsilon$, $\mu(x) < \varepsilon$— значение x с $\mu(x) \leq \varepsilon$ уже практически не принадлежит рассматриваемому нечёткому множеству.

При этом сами числа ε и λ являются порогами, используемыми в процессе построения шеститочесного представления нечёткого числа. Таким образом, шеститочечное представление нечёткого числа — это кусочно-линейная функция с заданными параметрами ε и λ:

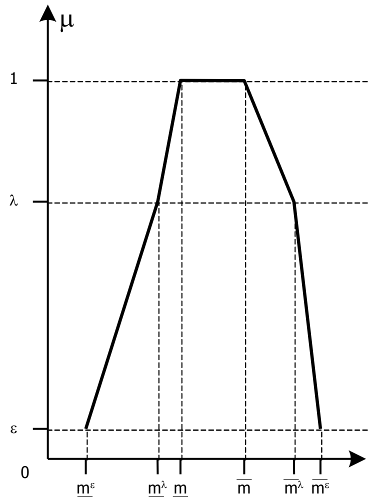

Рисунок 10. Шеститочечное нечёткое число

Из этого рисунка видно, что рассмотренные в предыдущем разделе триангулярные и трапециевидные нечёткие числа могут быть смоделированы шеститочечным представлением, так как такое представление может быть как унимодальным, так и толерантным. Кроме того, шеститочечное представление нечётких чисел является более широким, чем триангулярные и трапециевидные нечёткие числа. Поэтому класс задач, где можно найти применение этому формализму, несколько шире.

2.1.2. Кусочно-линейные функции принадлежности

Кусочно-линейное представление функций принадлежности нечётких множеств основано на аппроксимации гладких непрерывных функций отрезками прямых [19]. Формально такую аппроксимацию можно описать при помощи выбора ограниченного набора точек на области определения аппроксимируемой функции и восстановления функции между ними при помощи отрезков прямых. Таким образом, кусочно-линейная функция принадлежности может быть описана при помощи множества пар:

$$MF = \left\{\left\langle x_i, \mu(x_i)\right\rangle\right\}_{i=1}^{n},$$

где n — количество выбранных точек на интервале аппроксимации. Все промежуточные значения кусочно-линейной функции принадлежности в интервале между некоторыми x_1 и x_2 вычисляются посредством следующей формулы:

$$\mu(x) = \mu(x_1) + \frac{x - x_1}{x_2 - x_1}\left(\mu(x_2) - \mu(x_1)\right).$$

Такими кусочно-линейными функциями можно аппроксимировать функции принадлежности любой сложности без видимой потери точности. Варьирую количество точек аппроксимации можно добиваться большей или меньшей точности в представлении заданной функции. С другой стороны, такое представление позволяет оперировать функциями принадлежности без выполнения громоздких вычислений. Например, для построения пересечения двух кусочно-линейных функция принадлежности с количеством точек аппроксимации m и n соответственно необходимо проделать не более 2(m + n) вычислений, то есть задача имеет линейную сложность.

Следует отметить, что все описанные ранее способы представления нечётких чисел LR-типа (триангулярное, трапециевидное и шеститочечное) сами являются кусочно-линейными представлениями нечёткости, поэтому к таким формализмам можно применять общие алгоритмы вычислений различных операций над функциями принадлежности в кусочно-линейном представлении.

2.2. Представление неопределённости

Как уже было показано в разделе об извлечении знаний с неопределённостью, этот НЕ-фактор представим либо в виде конкретного числа из интервала [0; 1], либо непрерывным подинтервалом из этого же интервала. Все остальные способы представления неопределённых знаний являются теми или иными расширениями такого представления, не меняющими самой сути.

Таким образом, для представления неопределённости одного конкретного высказывания необходимо выделить память для содержания одного или двух чисел в зависимости от выбранного метода обработки неопределённых знаний. Например, метод Байеса предполагает хранение одного числа, а метод Демпстера-Шейфера — двух, нижней и верхней границ интервала неопределённости.

В работе [10] предлагается довольно необычный подход к представлению неопределённых знаний — *алгебраическая байесовская сеть*, которая является редуцированным вариантом вероятностного пространства. Показано, что модель алгебраической байесовской сети требует меньшего объёма исходной информации и позволяет строить эффективные алгоритмы оценивания и поддержания непротиворечивости знаний с неопределённостью.

Расширенное вероятностное пространство есть тройка:

$$E = \langle S, X, \mu \rangle,$$

где S — множество, $X = \{< X_i, \mu_i >\}_{i=1}^{N}$ — семейство множеств с вероятностной мерой таких, что $\forall i : X_i \subset S$, а μ_i — вероятностная мера X_i. При этом μ — вероятностная мера, определённая на алгебре семейства базовых множеств X.

В качестве примера можно рассмотреть задачу поиска объекта на поверхности Земли. Пусть геометрические соотношения между зоной локализации объекта и зонами, доступными для различных средств поиска, имеют вид, представленный на следующем рисунке:

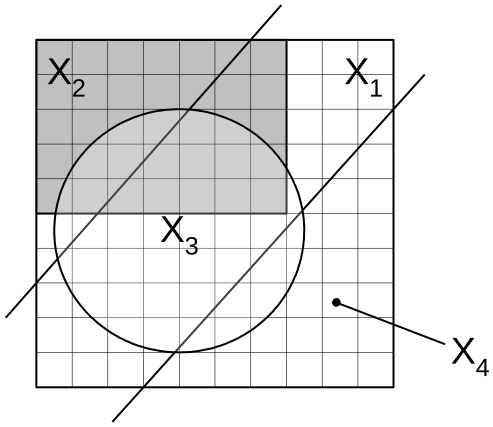

Рисунок 11. Области локализации объекта и зон видимости средств поиска

Здесь X_1 — зона наблюдения космического аппарата, X_2 — область Земли, попавшая на снимок с самолёта, X_3 — область «радиовидимости объекта», X_4 — оставшаяся область. Пусть мера множеств на приведённом рисунке задаётся множеством элементарных прямоугольников. Тогда области X_1, X_3 и X_4 неизмеримы в этом пространстве, а область X_2 — измерима, при этом оценки вероятностных мер множеств, полученные из геометрических отношений, таковы:

$$\begin{aligned} \mu(X_1) &\in [0.45,\ 0.60] \\ \mu(X_2) &\in [0.35,\ 0.35] \\ \mu(X_3) &\in [0.30,\ 0.48] \\ \mu(X_4) &\in [0.19,\ 0.31] \end{aligned}.$$

В этом примере роль множества S играет большой прямоугольник, система базовых подмножеств — это $X = \{< X_i, \mu_i >\}_{i=1}^{4}$, а значения вероятностных мер представлены выше. Сама мера задаётся системой малых прямоугольников, сумма которых равна 1, а мера каждого — 0.01.

Расширением этого пространства является система, в которой строится вероятностная логика $\langle S, X, \pi, \mu \rangle$, где

$$\begin{gathered} \pi = \{\pi_i\}, \pi_i : x_i \leftrightarrow X_i, x_i \in \Phi_0 = \{x_1, ..., x_n\} \\ x_i \in \{false,\ true\},\ X_i \in S, \bigcup_{i=1}^{n} X_i \subseteq S \end{gathered}.$$

Пусть Ф — множество правильно построенных формул над множеством пропозиций Φ_0. Каждой формуле $f \in \Phi$ ставится в соответствие вероятностная мера по стандартному изоморфизму алгебры пропозиционных формул и алгебры множеств $\{X_i\}$ по приведённому выше отображению:

$$\mu(\pi_i) : \mu(x_i) \leftrightarrow \mu(S_i),$$

при котором логические операции $\{\vee, \wedge, \neg\}$ соответствуют операциям $\{\cup, \cap, \overline{\ \ }\}$, а любой формуле из множества Ф ставится в соответствие вероятностная мера, приписанная её изоморфному образу в исходной системе базовых множеств. Такая логика называется логикой рассуждений о вероятностях. Построенная формальная система позволяет корректно описывать многие задачи с интервальными вероятностями.

Пусть задано расширенное вероятностное пространство $\langle S, X, \pi, \mu \rangle$. Предполагается, что в нём заданы только интервальные вероятностные меры некоторых формул из Φ. Такая ситуация соответствует тому, с чем реально приходится работать инженерам по знаниям. В частности, рассмотренный ранее пример именно таков. Формальная модель для такого случая может быть представлена в виде $\{\Phi_0, \langle f_i, \mu_i \rangle\}$, где $\langle f_i, \mu_i \rangle$ — множество формул с вероятностной мерой.

Вводятся дополнительные ограничения на вид построенной модели:

- высказывания экспертов содержат не более 2-3 пропозиций (обычно эксперт не в состоянии указать более сложные связи);
- все формулы в модели $\{\Phi_0, \langle f_i, \mu_i \rangle\}$ имеют вид конъюнкций, связывающих пропозиции без знаков отрицания.

Пусть $X = \{x_{i_1}, ..., x_{i_k}\} \subseteq \Phi_0$, $k \le n$ — множество элементарных пропозиций и K — множество всех возможных конъюнкций длины не более k, составленных из символов множества X. Пусть на множестве K введён частичный порядок (≤). Фрагментом знаний порядка k называется частично упорядоченное множество $A = \{\langle K, \mu \rangle, \le\}$, где μ — множество вероятностных мер, сопоставленных элементам множества K.

Множество $\{A_s\}_{s=1}^{N}$ фрагментов знаний с определённой непротиворечивым образом вероятностной мерой называется алгебраической байесовской сетью. Такая модель является упрощённым представлением расширенного вероятностного пространства, которое является удобной структурой для представления знаний с неопределённостью и удобна для решения задачи поддержки непротиворечивости баз знаний.

2.3. Представление неточности и недоопределённости

Как уже показано в определениях неточности и недоопределённости (см. Глоссарий), некоторые виды этих НЕ-факторов можно легко преобразовать друг к другу. Речь идёт о так называемых интервальных НЕ-факторах. С другой стороны другие виды неточности и недоопределённости пока ещё мало изучены [21, 22].

Представление интервальных НЕ-факторов подразумевается самим названием — ответ на вопрос «Как представлять?» довольно прост: представлять при помощи интервалов на действительной оси **R**. Но с другой стороны второй вопрос представления знаний остаётся открытым, так как для двух рассматриваемых НЕ-факторов ещё не разработаны математические аппараты, полностью охватывающие все аспекты применения неточности и недоопределённости.

Одним из видов неточности является методическая неточность, которая проявляется вместе с числовыми значениями, полученными при помощи каких-либо измерительных приборов с заранее оговорённой погрешностью измерения. Такие неточные значения довольно легко представимы при помощи двух действительных чисел:

$$x = x_0 \pm \Delta x,$$

где x_0 — измеренное значение величины x, а Δx — абсолютная погрешность измерительного прибора, при помощи которого измерена величина x_0. Обычно на приборах указана относительная погрешность измерения, которую можно преобразовать в абсолютную и обратно по следующим формулам:

$$\delta x = \frac{\Delta x}{x_0} \quad .$$
$$\Delta x = x_0 \cdot \delta x$$

Недоопределённость проявляется тогда, когда для какой-либо величины невозможно определённо установить её значение, однако в процессе решения задачи это значение постепенно приближается к истинному (более определённому), то есть значение как бы доопределяется. Такое недоопределённое значение находится внутри какого-либо интервала недоопределённости, который изменяется в процессе поиска истинного значения исследуемой величины. Такое изменение может происходить как в сторону доопределения, так и в обратную сторону в зависимости от конкретной задачи и проблемной области. Кроме того, интервальная арифметика предполагает, что некоторые величины могут находиться не в одном интервале, а сразу в некотором множестве непересекающихся интервалов (см. далее раздел об интервальной арифметике).

Необходимо различать неточность и недоопределённость, так как эти НЕ-факторы проявляются в решаемых задачах по-разному. Если недоопределённое значение какой-либо величины необходимо как бы доопределять в процессе решения задачи, то в повышении точности неточной величины, полученной при помощи измерительного прибора, зачастую нет особого смысла. Например, точность в 1 мм бессмысленна для решения задачи, где фигурируют расстояния порядка метров.

Таким образом, представление неточности и недоопределённости можно производить при помощи интервалов на действительной оси. Методическая неточность всегда представляется в виде одного интервала, в то время как недоопределённость знаний может в некоторых случаях быть представлена при помощи множества непересекающихся интервалов. То есть интервальные НЕ-факторы могут быть представлены в виде множества:

$$I = \{[a_i; b_i]\}_{i=1}^{n},$$

где n — количество интервалов, которое для большинства случаев равно единице.

При этом следует помнить, что такое множество должно состоять из упорядоченных и непересекающихся интервалов на оси действительных чисел **R**, то есть:

$$\forall i < n : b_i < a_{i+1}.$$

Знак «строго меньше» в предыдущей формуле стоит по той причине, что если $b_i = a_{i+1}$, то в этом случае нет необходимости хранить два интервала и можно обойтись одним, так как границы интервалов совпадают, и их можно «слить» в один.

Глава 3. Обработка знаний с НЕ-факторами

В настоящей монографии обработка знаний рассматривается с двух точек зрения — подготовка знаний к выводу, то есть верификация приобретённых и извлечённых знаний из источников знаний, а также собственно вывод на продукциях. Хотя обычно верификация предшествует выводу, в дальнейшем изложении сначала рассматриваются методы машинного вывода на знаниях с НЕ-факторами, а только потом изучаются вопросы верификации информации с НЕ-факторами, полученной с этапов приобретения и извлечения знаний. Кроме того, отдельным разделом стоит описание методов преобразования НЕ-факторов, то есть перевода из одного вида в другой, когда это возможно.

3.1. Вывод на знаниях с НЕ-факторами

В этом разделе будет рассматриваться только машинный вывод на продукциях. По возможности рассматриваются различные технологии и методы прямого, обратного и смешанного продукционного вывода для четырёх выделенных НЕ-факторов: нечёткости, неопределённости, неточности и недоопределённости. В случаях, где это необходимо, приводятся теоретические основы математического аппарата, использованного в том или ином подходе.

3.1.1. Вывод на нечётких знаниях

Как уже указывалось, нечёткая логика — это бесконечнозначная логика, значения истинности которой лежат в интервале [0; 1]. Это является причиной некоторой вольности, которая выражается в том, что операции с нечёткими величинами не определены однозначно, как в логике Аристотелевой, а могут выбираться из целых классов соответствующих операций [26]. Такими классами являются треугольные нормы и треугольные конормы, которые определяют операции пересечения и объединения соответственно.

Треугольной нормой (Т-нормой) называется двухместная действительная функция $T : [0; 1] \times [0; 1] \to [0; 1]$, удовлетворяющая следующим условиям:

1. Ограниченность: $T(0, 0) = 0$; $T(\mu_A, 1) = \mu_A$; $T(1, \mu_A) = \mu_A$.
2. Монотонность: $T(\mu_A, \mu_B) \leq T(\mu_C, \mu_D)$, если $\mu_A \leq \mu_C$ и $\mu_B \leq \mu_D$.
3. Коммутативность: $T(\mu_A, \mu_B) = T(\mu_B, \mu_A)$.
4. Ассоциативность: $T(\mu_A, T(\mu_B, \mu_C)) = T(T(\mu_A, \mu_B), \mu_C)$.

Треугольной конормой (Т-конормой) называется двухместная действительная функция $S : [0; 1] \times [0; 1] \to [0; 1]$, удовлетворяющая следующим условиям:

1. Ограниченность: $S(1, 1) = 1$; $S(\mu_A, 0) = \mu_A$; $S(0, \mu_A) = \mu_A$.
2. Монотонность: $S(\mu_A, \mu_B) \geq S(\mu_C, \mu_D)$, если $\mu_A \geq \mu_C$ и $\mu_B \geq \mu_D$.
3. Коммутативность: $S(\mu_A, \mu_B) = S(\mu_B, \mu_A)$.
4. Ассоциативность: $S(\mu_A, S(\mu_B, \mu_C)) = S(S(\mu_A, \mu_B), \mu_C)$.

В качестве примеров треугольных норм и конорм можно привести следующие операции:

Таблица 6. Примеры нечётких операций из класса Т-норм и Т-конорм

№	Т-норма	Т-конорма
1	$\min(\mu_A, \mu_B)$	$\max(\mu_A, \mu_B)$
2	$\mu_A \cdot \mu_B$	$\mu_A + \mu_B - \mu_A \cdot \mu_B$
3	$\max(0, \mu_A + \mu_B - 1)$	$\min(1, \mu_A + \mu_B)$

Представленные нечёткие операции далеко не единственные, которые можно применять в машинном выводе. Более того, существуют классы параметрических операций (например, параметрические Т-нормы и Т-конормы Сугено [26]), зависящие от определённого параметра и позволяющие тонко настраивать механизм вывода. Однако обобщённая схема нечёткого вывода определена одинаково для всех операций из классов Т-норм и Т-конорм.

Продукционный нечёткий вывод предполагает, что описание знаний о проблемной области сформулировано экспертами в виде набора правил вида:

$$
\begin{gathered}
\text{Rule}_1\text{: Если } x_1 \text{ есть } A_1\text{, то } y_1 \text{ есть } B_1\text{,} \\
\text{Rule}_2\text{: Если } x_2 \text{ есть } A_2\text{, то } y_2 \text{ есть } B_2\text{,} \\
\ldots \\
\text{Rule}_n\text{: Если } x_n \text{ есть } A_n\text{, то } y_n \text{ есть } B_n\text{.}
\end{gathered}
$$

где X — множество имён входных переменных, Y — множество имён переменных вывода, а A_i и B_i — функции принадлежности (имена функций принадлежности), определённые для x и y соответственно. Примером правила подобного вида может выступать следующее:

Если distance = «близко», то height = «высоко».

Примечание: указанное правило оторвано от какой-либо проблемной области и приведено только в целях написания примера.

К указанному набору продукций в процессе нечёткого вывода применяют модифицированные правила Modus Ponens и Modus Tollens. Сами модифицированные правила выглядят следующим образом [12]:

Modus Ponens:

$$
\begin{array}{l}
\text{Если x есть A, то y есть B} \\
\text{x есть A'} \\
\hline
\text{y есть B'}
\end{array}
$$

Modus Tollens:

$$
\begin{array}{l}
\text{Если x есть A, то y есть B} \\
\text{y есть B'} \\
\hline
\text{x есть A'}
\end{array}
$$

В соответствии с принципом резолюций [33, 27] и свойствами нечётких операций, модифицированные правила Modus Ponens и Modus Tollens можно записать в виде уравнений [23, 18]:

$$
\begin{gathered}
B' = A' \bullet (A \to B) \\
A' = B' \bullet (A \to B)
\end{gathered},
$$

где знак «•» обозначает операцию композиции или свёртки, которая определяется как:

$$\mu_{A\bullet B}(x,z) = \vee_y (\mu_A(x,y) \wedge \mu_B(y,z)).$$

Естественно, что как операция композиции, так и операция импликации может быть определена в нечёткой математике различными способами, поэтому будет разниться и получаемый в итоге вывода результат, однако в любом случае общая методика продукционного вывода на нечётких знаниях осуществляется согласно следующему списку этапов [18, 26]:

1. *Фаззификация*.

Для каждого входного фактического параметра (переменной, участвующей в процессе вывода), определяется соответствующая функция принадлежности. Для этого применяются синтаксические и семантические процедуры той лингвистической переменной, значением которой является текущий входной параметр.

2. *Сопоставление*.

Для посылки (антецедента) каждого правила, участвующего на очередном шаге вывода, вычисляется значение истинности, которое в дальнейшем применяется к заключению (консеквенту) правила. Это приводит к некоторому видоизменению функции принадлежности в консеквенте правил, и это видоизменение зависит от используемого метода вывода.

3. *Композиция*.

Все функции принадлежности, полученные в процессе сопоставления и относящиеся к одной и той же переменной вывода, объединяются для того, чтобы сформировать одну функцию принадлежности. Способ композиции так же зависит от используемого метода нечёткого вывода.

4. *Дефаззификация* (опционально).

Дефаззификация — это приведение функций принадлежности к чётким значениям. Эта процедура используется тогда, когда полезно преобразовать набор выведенных значений, представляющий собой множество функций принадлежности, в чёткие значения, которые можно интерпретировать в терминах проблемной области. Эта процедура в свою очередь не зависит от метода нечёткого вывода, однако сама может варьироваться, поэтому в процессе нечёткого вывода целесообразно иметь механизм выбора способа дефаззификации.

3.1.1.1. Прямой нечёткий вывод

Все алгоритмы прямого нечёткого вывода целесообразно рассматривать на примере двух правил:

Rule$_1$: Если x есть A_1 **И** y есть B_1, то z есть C_1
Rule$_2$: Если x есть A_2 **И** y есть B_2, то z есть C_2

Здесь x и y — имена входных переменных, z — имя переменной вывода. A_1, A_2, B_1, B_2, C_1 и C_2 — некоторые заданные функции принадлежности. Кроме того, заданы входные (фактические) значения переменных x и y — x_0 и y_0 соответственно. Эти значения являются чёткими, поэтому перед процессом вывода их необходимо фаззифицировать. По этим входным значениям необходимо получить чёткое выходное значение z_0, поэтому в конце вывода требуется осуществить дефаззификацию.

Необходимо отметить, что для всех приведённых далее алгоритмов прямого нечёткого вывода процесс фаззификации осуществляется абсолютно одинаково, поэтому для экономии места целесообразно описать его здесь: для обоих правил находятся степени истинности каждого выражения в антецеденте — $A_1(x_0)$, $A_2(x_0)$, $B_1(y_0)$ и $B_2(y_0)$. С другой стороны, для каждого чёткого значения x_0 и y_0 можно полагать существующей соответствующую функцию принадлежности, и эти функции принадлежности определяются следующим образом:

$$F_x(x_0) = 1; \forall x \neq x_0 : F_x(x) = 0$$
$$F_y(y_0) = 1; \forall y \neq y_0 : F_y(y) = 0.$$

В этом случае процесс нахождения степеней истинности для чётких значений переменных x и y сводится к нахождению пересечения функций принадлежности, которое, как уже было замечено, не определено однозначно, а может выбираться из класса T-норм. В большинстве случаев для вычисления пересечения функций на данном этапе используют операцию min [23].

В общем виде прямой нечёткий вывод ничем не отличается от обычного продукционного вывода, кроме добавления двух этапов — фаззификации и дефаззификации:

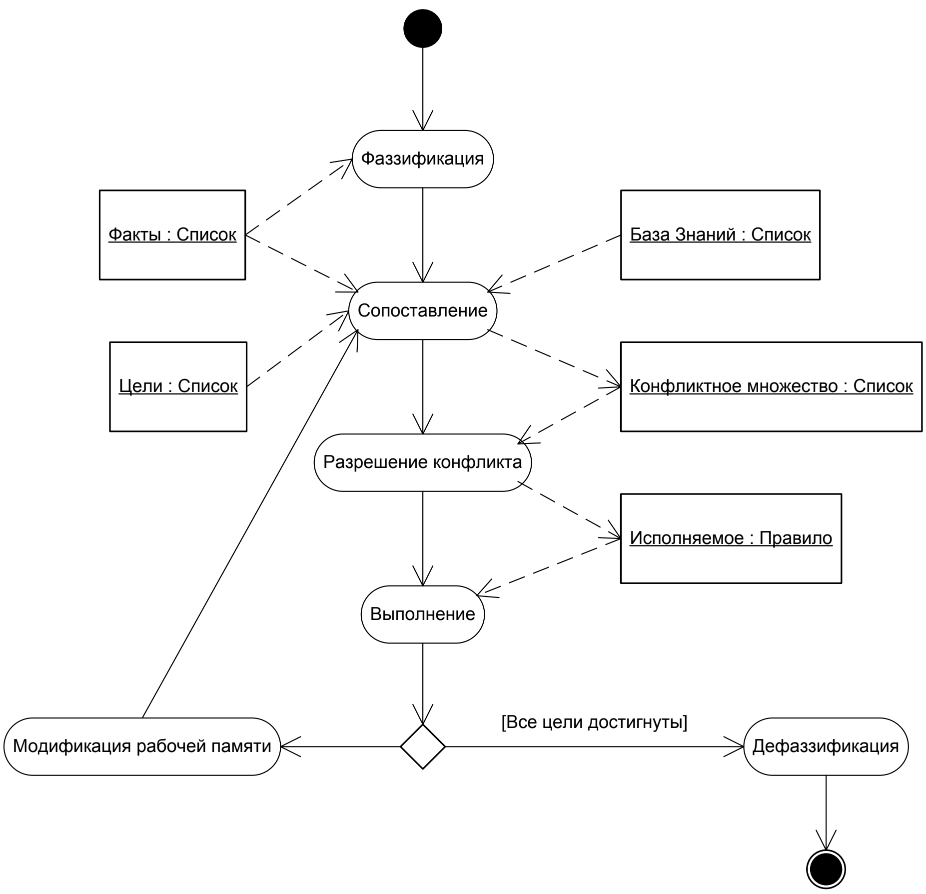

Рисунок 12. Общая схема прямого нечёткого вывода

Как видно из представленного рисунка дополнительные шаги в процессе прямого нечёткого вывода выполняются в самом начале (фаззификация) и в самом конце (дефаззификация) процесса. Остальные шаги остаются без изменения.

3.1.1.1.1. Алгоритм Mamdani

Уровни усечения в процессе сопоставления находятся при помощи применения операции min:

$$\begin{aligned}\alpha_1 &= \min(A_1(x_0), B_1(y_0)) \\ \alpha_2 &= \min(A_2(x_0), B_2(y_0))\end{aligned}.$$

После этого функции принадлежности консеквентов правил усекаются по найденным значениям истинности:

$$\begin{aligned}C_1^{'} &= \min(\alpha_1, C_1) \\ C_2^{'} &= \min(\alpha_2, C_2)\end{aligned}.$$

Композиция осуществляется при помощи операции max, то есть для всех целевых переменных объединяются найденные в процессе вывода функции принадлежности (уже усечённые). В рассматриваемом примере конечная функция C(z), которая будет подана на вход процедуры дефаззификации, вычисляется следующим образом:

$$C(z) = \max(C_1^{'}, C_2^{'}) = \max(\min(\alpha_1, C_1), \min(\alpha_2, C_2)).$$

Для дефаззификации обычно используется метод центра тяжести (см. далее в разделе «Метод центра тяжести»).

На следующем рисунке показана иллюстрация к прямому нечёткому выводу по алгоритму Mamdani. На рисунке не показана дефаззификация, а под графой «Результат» выделена полученная функция принадлежности C(z).

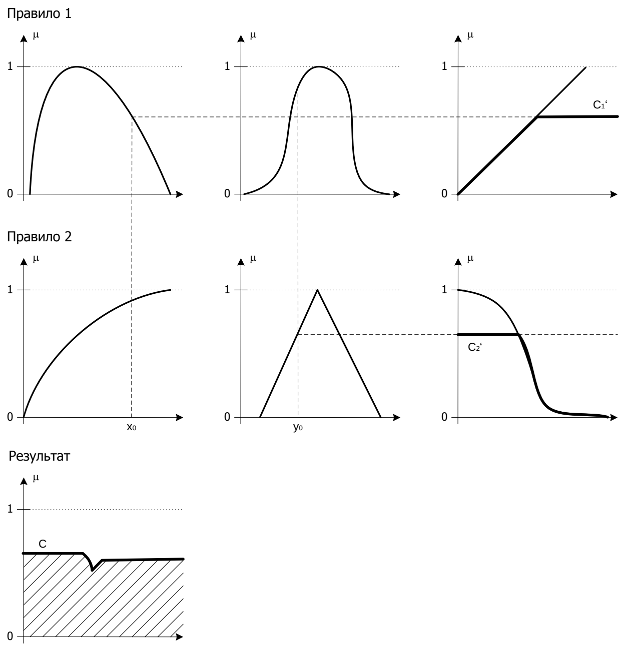

Рисунок 13. Процесс прямого нечёткого вывода по алгоритму Mamdani

3.1.1.1.2. Алгоритм Tsukamoto

Уровни усечения функций принадлежности консеквентов правил находятся абсолютно так же, как и в алгоритме Mamdani, однако затем для каждого правила находится чёткое значение при помощи решения уравнений следующего вида:

$$\begin{aligned}\alpha_1 &= C_1(z_1)\\ \alpha_2 &= C_2(z_2)\end{aligned}.$$

По найденным чётким значениям z_1 и z_2 определяется конечное значение целевой переменной по формуле взвешенного среднего:

$$z_0 = \frac{\alpha_1 z_1 + \alpha_2 z_2}{\alpha_1 + \alpha_2}.$$

Причём полученное значение z_0 уже представляет собой чёткое значение, поэтому выполнять процедуру дефаззификации не надо. Приведённую выше формулу взвешенного среднего естественно можно распространить на случай n правил.

На следующем рисунке показана иллюстрация к прямому нечёткому выводу по алгоритму Tsukamoto.

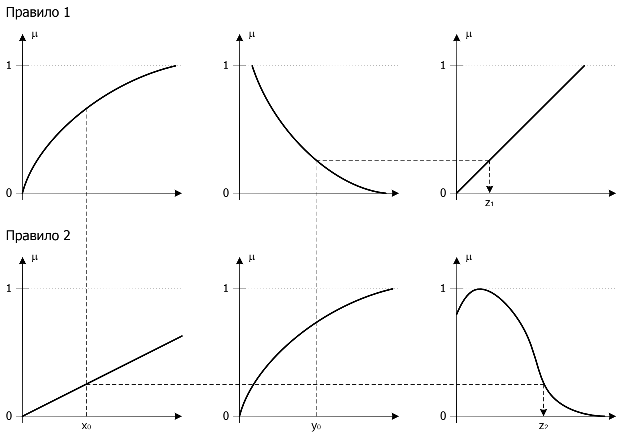

Рисунок 14. Процесс прямого нечёткого вывода по алгоритму Tsukamoto

3.1.1.1.3. Алгоритм Sugeno

В алгоритме Sugeno полагается, что правила в базе знаний имеют вид, отличающийся от приведённого ранее, а именно:

$Rule_1$: Если x есть A_1 **И** y есть B_1, то z есть $a_1x + b_1y$
$Rule_2$: Если x есть A_2 **И** y есть B_2, то z есть $a_2x + b_2y$

где a_1, a_2, b_1 и b_2 — суть некоторые чёткие коэффициенты.

В этом случае уровни усечения находятся по тем же формулам, что и в алгоритме Mamdani, однако затем для всех правил с целевыми переменными в консеквентах находятся частные выводы этих правил:

$$\begin{aligned}z_1^* &= a_1 x_0 + b_1 y_0\\ z_2^* &= a_2 x_0 + b_2 y_0\end{aligned}.$$

А значение (уже дефаззифицированное) так же как и в алгоритме Tsukamoto находится при помощи вычисления взвешенного среднего:

$$z_0 = \frac{\alpha_1 z_1^* + \alpha_2 z_2^*}{\alpha_1 + \alpha_2}.$$

3.1.1.1.4. Алгоритм Larsen

В алгоритме Larsen вычисление уровней усечения вычисляются также, как и в алгоритме Mamdani, и затем эти уровни используются для вычисления частных нечётких подмножеств консеквентов правил, участвующих в выводе. Частные функции принадлежности вычисляются при помощи умножения на соответствующий уровень усечения (в этом алгоритме уровни усечения целесообразно называть коэффициентами истинности):

$$\begin{matrix} C'_1 = \alpha_1 C_1 \\ C'_2 = \alpha_2 C_2 \end{matrix}.$$

После получения новых функций принадлежности выполняется композиция этих функций при помощи их объединения. Дефаззификацию можно производить любым из доступных методов.

На следующем рисунке проиллюстрирована работа алгоритма Larsen с двумя правилами, в посылке каждого из которых используется конкатенация двух нечётких высказываний. Под графой «Результат» показана конечная функция принадлежности, которую можно дефаззифицировать.

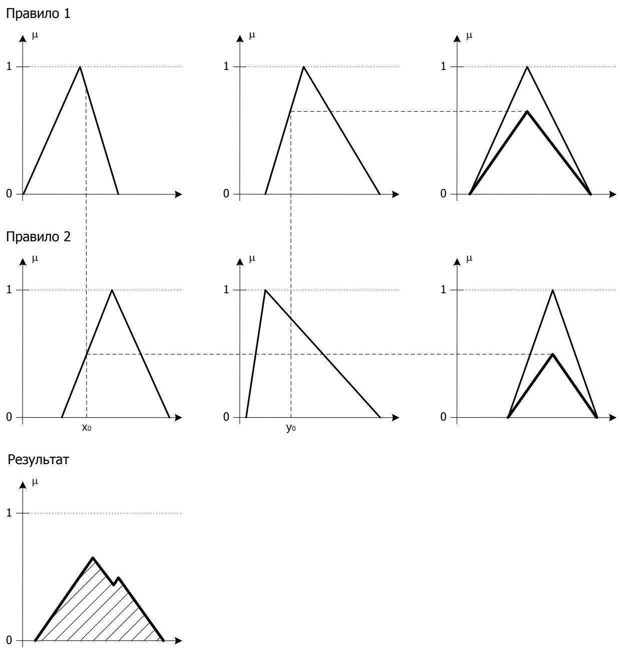

Рисунок 15. Процесс прямого нечёткого вывода по алгоритму Larsen

3.1.1.1.5. Упрощённый алгоритм прямого нечёткого вывода

Упрощённый алгоритм прямого нечёткого вывода используется в тех случаях, когда в консеквентах правил целевые переменные уже принимают чёткие значения, то есть правила принимают вид:

$Rule_1$: Если x есть A_1 **И** y есть B_1, то z есть c_1

$Rule_2$: Если x есть A_2 **И** y есть B_2, то z есть c_2

В этом случае можно применять алгоритм Tsukamoto, но без решения уравнений, связанных с нахождением частных чётких значений каждого правила, так как эти значения уже даны в консеквентах правил. Окончательное дефаззифицированное значение целевой переменной вычисляется по формуле взвешенного среднего.

3.1.1.1.6. Максиминный метод нечёткого вывода

Максиминный метод прямого нечёткого вывода [12, 26] также использует правила, общий вид которых показан в начале настоящего раздела. В своей сущности, этот метод обобщает алгоритм Mamdani до случая, когда в качестве фактических значений переменных в антецедентах правил могут выступать не только чёткие, но и нечёткие значения, определённые соответствующими функциями принадлежности.

В этом случае уровни усечения находятся при помощи все той же операции min, которая применяется к фактической функции принадлежности и функции принадлежности из антецедента правила:

$$\begin{aligned}\alpha_1 &= \min(\min(A_1, A'), \min(B_1, B')) \\ \alpha_2 &= \min(\min(A_2, A'), \min(B_2, B'))\end{aligned},$$

где A’ и B’ — функции принадлежности фактических значений переменных x и y.

Затем для оптимизации вычислений из всех уровней усечения выбирается максимальный, и все дальнейшие вычисления ведутся с правилом, которому соответствует выбранный уровень усечения. Дефаззификация в рассматриваемом методе может производиться любым методом, однако в большинстве случаев используется метод центра тяжести.

3.1.1.2. Обратный нечёткий вывод

Обратный вывод на продукциях отличается от прямого тем, что в самом начале процесса вывода определены значения не исходных фактов, а целевых переменных (заключений, симптомов), а в самом процессе вывода определяются значения посылок (входы, факторы).

Предполагается, что вся продукционная база знаний, содержащая правила с нечёткими параметрами, может быть представлена в виде матрицы нечётких отношений R, состоящая из элементов $r_{ij} \in [0; 1]$, при этом каждый элемент r_{ij} может быть найден из соотношения:

$$r_{ij} = x_i \rightarrow y_j.$$

То есть коэффициент r_{ij} — это нечёткое причинное отношение, вычисляемое для каждого правила в базе знаний. В случае если значениями переменных в правилах являются функции принадлежности, коэффициенты r_{ij} можно вычислять как максимумы пересечения соответствующих функций на своих областях определения [23].

Пояснение методики обратного нечёткого вывода целесообразно проводить на примере. Для этого достаточно взять очень упрощённую модель диагностики неисправности автомобиля [18]. Пусть в рассматриваемой модели существуют следующие параметры:

- x_1 — неисправность аккумулятора.
- x_2 — отработка машинного масла.
- y_1 — затруднение при запуске.
- y_2 — ухудшение цвета выхлопных газов.
- y_3 — недостаток мощности.

Пусть знания эксперта-автомеханика имеют следующий вид:

$$R = \begin{bmatrix} 0.9 & 0.1 & 0.2 \\ 0.6 & 0.5 & 0.5 \end{bmatrix}.$$

Эти коэффициенты обозначают, что, например, при неисправности аккумулятора эксперт-автомеханик с уверенностью в 90 % (0.9) может предположить, что произойдут затруднения при пуске. Остальные коэффициенты трактуются подобным образом.

Пусть в результате осмотре некоторого конкретного автомобиля его состояние оценивается следующим образом:

$$Y = 0.6 \mid y_1 + 0.1 \mid y_2 + 0.1 \mid y_3.$$

В процессе решения задачи следует определить причину такого состояния обследуемого автомобиля, то есть найти коэффициенты α_1 и α_2:

$$X = \alpha_1 \mid x_1 + \alpha_2 \mid x_2.$$

Для того чтобы решить поставленную задачу, достаточно решить следующую систему уравнений:

$$\begin{bmatrix} 0.6 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.6 \\ 0.1 & 0.5 \\ 0.2 & 0.5 \end{bmatrix} \bullet \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}.$$

Или, что то же:

$$\begin{array}{c} 0.6 = (0.9 \wedge \alpha_1) \vee (0.6 \wedge \alpha_2) \\ 0.1 = (0.1 \wedge \alpha_1) \vee (0.5 \wedge \alpha_2) \\ 0.1 = (0.2 \wedge \alpha_1) \vee (0.5 \wedge \alpha_2) \end{array}.$$

Соответственно операции $\wedge$ и $\vee$ — это Т-норма и Т-конорма соответственно. В большинстве случаев используются операции min и max.

Из первого уравнения можно заключить, что оба коэффициента α_1 и α_2 меньше значения 0.6. Из второго уравнения можно заключить, что второй коэффициент не может превышать значение 0.1. Наконец, третье уравнение ничего нового не даёт, поэтому окончательное решение будет выглядеть следующим образом:

$$\begin{array}{c} 0.0 \le \alpha_1 \le 0.6 \\ 0.0 \le \alpha_2 \le 0.1 \end{array}.$$

Полученные результаты можно трактовать различными способами. Проще всего полученные коэффициенты трактовать как уверенность в посылках. То есть в рассматриваемом случае можно с 60 % уверенностью утверждать, что подсел аккумулятор, и только с 10 % уверенностью предположить, что следует заменить масло в картере двигателя.

На практике в задачах, подобных рассмотренной, количество переменных может быть исключительно большим, кроме того, могут одновременно использоваться различные композиции, Т-нормы и Т-конормы для решения систем уравнений, а сама стратегия вывода может включать в себя и «прямые участки». По всей видимости, общих методов решения таких обобщённых задач в настоящее время не существует [18].

3.1.1.3. Нечёткий вывод, использующий преобразование функций принадлежности

Особенный метод нечёткого вывода был предложен в работе [7]. Этот метод основан на преобразовании функций принадлежности и используется на практике в управлении сложными системами. Методы управления, основанные на опыте и знаниях эксперта, предполагают, что на объект управления, имеющий нечёткие состояния, производятся воздействия в зависимости от нечётких лингвистических правил.

Предлагаемый метод относится к классу «методов нечёткого вывода с преобразованием функций принадлежности», при этом в нём отсутствует понятие нечёткого отношения. Это позволяет работать с более широким классом функций принадлежности и производить более качественный вывод на нечётких множествах.

В предлагаемом методе нечёткая продукция $A \Rightarrow B$ рассматривается, как отображение нечёткой величины A (функция принадлежности $\mu_A(x)$) в нечёткую величину B (функция принадлежности $\mu_B(x)$). Задача заключается в нахождении нечёткой величины B_1 при наличии нечёткой величины A_1 (полученной в процессе фаззификации) и правила преобразования $A \Rightarrow B$.

3.1.1.3.1. Ограничения на вид функций принадлежности

Функция принадлежности должна обладать некоторыми важными характеристиками. Для описания рассматриваемого метода нечёткого вывода необходимо задать класс функций, описывающих функции принадлежности нечётких величин. При выборе достаточно широкого класса функций любое нечёткое множество можно аппроксимировать функцией из этого класса без потери точности представления.

Пусть диапазон изменения аргумента функции принадлежности есть $[h_1; h_2]$. Для рассмотрения метода этот диапазон необходимо нормировать в интервал [0; 1], то есть получаемые таким образом функции принадлежности имеют общий вид:

$$\mu(x) : [0;1] \to [0;1].$$

Кроме того, функция принадлежности должна быть непрерывной, везде определённой (на своей области определения), а также её производная не должна равняться нулю (отсутствие экстремумов).

3.1.1.3.2. Методы преобразования функций принадлежности

Далее рассматриваются три метода нечёткого вывода с использованием преобразования функции принадлежности. В основу каждого из этих методов положены различные предположения о характере трансформации функции принадлежности. Класс используемых функций принадлежности можно расширять для отдельных методов, а использование параметрических функций принадлежности увеличивает возможности рассматриваемых методов. Подбор параметрических семейств функций принадлежности возможен при конкретизации решаемой задачи.

Метод операторного преобразования

Пусть существует линейный оператор F, переводящий функцию $\mu_A : [0;1] \to [0;1]$ в функцию μ_B, определяемую также. Таким образом, линейный оператор F отображает единичный квадрат [0; 1] × [0; 1] в самого себя. Оператор ставит каждой функции в соответствие другую функцию из того же множества [0; 1] × [0; 1]. Пусть этот факт записывается как:

$$F(\mu_A) = \mu_B.$$

Таким образом, поиск неизвестного нечёткого множества B_1 сводится к решению уравнения:

$$\mu_{B_1} = F(\mu_{A_1}).$$

Для решения этого уравнения его необходимо продифференцировать, равно как необходимо продифференцировать и определение линейного оператора F:

$$\begin{gathered} \frac{dF(\mu_A)}{d\mu_A}\frac{d\mu_A}{dx} = \frac{d\mu_B}{dx} \\ \frac{dF(\mu_{A_1})}{d\mu_{A_1}}\frac{d\mu_{A_1}}{dx} = \frac{d\mu_{B_1}}{dx} \end{gathered}.$$

Главный принцип рассматриваемого метода заключается в следующем равенстве:

$$\frac{dF(\mu_A)}{d\mu_A} = \frac{dF(\mu_{A_1})}{d\mu_{A_1}} = \frac{dF(z)}{dz},$$

где z — любая функция из подпространства действительных чисел **R**, то есть $z : [0;1] \to [0;1]$. То есть производная линейного оператора F по аргументу одинакова для каждого нечёткого вывода и не зависит от самого аргумента. Учитывая этот факт, можно получить следующие результаты:

$$\frac{d\mu_{B_1}}{dx} = \frac{\frac{d\mu_B}{dx}\frac{d\mu_{A_1}}{dx}}{\frac{d\mu_A}{dx}}.$$

После интегрирования последнего выражения и осуществления подстановки $\frac{d\mu}{dx} = \dot{\mu}$ имеет место:

$$\mu_{B_1}(x) = \int_0^x \frac{\dot{\mu}_B \dot{\mu}_{A_1}}{\dot{\mu}_A} dx.$$

Эта формула неприменима в случае, если $\dot{\mu}_A = 0$, то есть в тех точках, где функция принадлежности имеет экстремум. Если метод применяется к таким функциям, то необходимо эти функции разбить на интервалы монотонности. На каждом из таких интервалов восстанавливается функция принадлежности μ_{B_1}. В точках разрыва искомая функция сшивается по принципу непрерывности.

При интегрировании искомая функция может выйти за пределы единичного интервала [0; 1]. В этом случае после интегрирования необходимо произвести нормализацию.

Метод операторной интерполяции

В задачах управления сложными объектами может возникнуть ситуация, когда в нечётком выводе применимы несколько нечётких правил. Это возможно, например, когда значение управляющей переменной оказывается на стыке нескольких функций принадлежности, соответствующих различным нечётким правилам. Выводимое заключение должно согласовываться с каждым из правил. В предыдущем методе функция принадлежности B_1 вычислялась с использованием только одного правила $A \Rightarrow B$. То есть оператор преобразования строился по одной точке (μ_A; μ_B) в пространстве функций принадлежности. Метод операторной интерполяции позволяет строить неизвестную функцию принадлежности по набору нечётких правил вывода, что существенно повышает точность необходимого нечёткого вывода.

Пусть существует набор правил вида $A_i \Rightarrow B_i$, где $i = 1,n$ с функциями принадлежности μ_{A_i} и μ_{B_i} соответственно. Пусть также существует множество A_{n+1}, которому необходимо поставить в соответствие нечёткое множество B_{n+1}. Необходимо отметить, что все $A_i \Rightarrow B_i$ применимы для решения поставленной задачи.

Для решения этой проблемы необходимо воспользоваться интерполяционным многочленом Лагранжа. На данные n + 1 точек в пространстве функций, по теореме единственности, можно наложить только один многочлен степени n. Пусть есть несовпадающие точки $x_1, ..., x_n$ для которых известны значения функций $F(x_1), ..., F(x_n)$. Тогда в рассматриваемом случае интерполяционная формула Лагранжа имеет вид:

$$F(x_{n+1}) = \sum_{k=0}^{n} \frac{\prod_{i=0, i \neq k}^{n} (x_{n+1} - x_i)}{\prod_{i=0, i \neq k}^{n} (x_k - x_i)} F(x_k).$$

Метод Лагранжа предназначен для построения функции в пространстве значений переменных. В случае рассматриваемого метода строится функционал в пространстве функций. Этот факт не накладывает дополнительных ограничений. Необходимо обратить внимание на то, что разность функций предпосылок нечётких правил не должна равняться нулю. В пространстве функций принадлежности посылок и заключений предполагаются заданными несколько точек. Используя эти точки, можно задать траекторию в операторном пространстве, и тем самым аппроксимировать неизвестную функцию принадлежности.

Считая, что интерполяционный многочлен Лагранжа позволяет вычислить искомую функцию принадлежности, можно получить формулу:

$$\mu_{B_{n+1}} = \sum_{k=0}^{n} \frac{\prod_{i=0, i \neq k}^{n} (\mu_{A_{n+1}} - \mu_{A_i})}{\prod_{i=0, i \neq k}^{n} (\mu_{A_k} - \mu_{A_i})} \mu_{B_k}.$$

Эта формула предполагает, что $\mu_{A_i} \neq \mu_{A_k}$ при $i \neq k$. Если существуют такие i и k, при которых функции принадлежности посылок равны, то такой случай необходимо рассматривать отдельно, так как в этом случае формула неприменима — знаменатель обращается в ноль. Такая ситуация возможна в случае противоречивости базы знаний. Однако рассмотрение этого случая не входит в задачи данного обзора.

Метод пропорциональности

В основу этого метода положен принцип, когда отдельно рассматривается каждое значение x на всех четырёх графиках функций принадлежности нечётких множеств A, B, A_1 и B_1. В случае, когда $A \Rightarrow B$ и $A_1 \Rightarrow B_1$, график μ_A непрерывно отображается в график μ_B. Такой же принцип непрерывного отображения должен быть и для функции μ_{A_1}. Пусть нечёткий вывод таков, что существует некоторая функция f(x), такая, что

$$\mu_A(x) f(x) = \mu_B(x).$$

Для каждого значения x отношение функций принадлежности в этом случае будет константой. То есть:

$$\frac{\mu_A}{\mu_B} = \frac{\mu_{A_1}}{\mu_{B_1}} = const,$$

откуда

$$\mu_{B_1} = \frac{\mu_{A_1} \mu_B}{\mu_A}.$$

Для этого метода необходимо, чтобы $\mu_A \neq 0$. Это есть частный случай метода операторного преобразования. Для него характерна линейная зависимость в пространстве функций.

3.1.1.4. Методы дефаззификации

Как уже говорилось, дефаззификация — это процесс приведения нечётких значений (описываемых функциями принадлежности) к чётким. Все методы дефаззификации сводятся к выбору одной конкретной точки на области определения дефаззифицируемой функции принадлежности.

Для каждого метода дефаззификации по возможности будет описан непрерывный и дискретный вариант, так как функции принадлежности некоторых нечётких множеств в свою очередь могут быть непрерывными (например, в случае представления в виде сигмоид или гауссиан) и дискретными (например, в случае представления кусочно-линейными функциями).

3.1.1.4.1. Метод центра тяжести

Метод центра тяжести основан на вычислении центра тяжести функции, и получаемое значение объявляется дефаззифицированным значением функции принадлежности. Для непрерывных функций принадлежности формула основана на интегрировании:

$$z_0 = \frac{\int_{\Omega} zC(z)dz}{\int_{\Omega} C(z)dz}.$$

Тот же вариант для дискретных функций принадлежности:

$$z_0 = \frac{\sum_{i=1}^{n} \alpha_i z_i}{\sum_{i=1}^{n} \alpha_i}.$$

Иногда возникают ситуации, когда метод центра тяжести даёт на области определения дефаззифицируемой функции принадлежности точку, которая имеет самую низкую степень принадлежности рассматриваемому нечёткому множеству. Например, если дефаззифицировать нечёткое отрицание нормального треугольного нечёткого числа с ограниченной областью определения и симметричного относительно точки максимума, то результатом дефаззификации будет чёткое значение, значение функции принадлежности в котором — 0.

Конечно, функции принадлежности с несколькими локальными максимумами в процессе приобретения знаний из экспертов встречаются крайне редко. Однако в процессе решения реальных задач такие функции могут возникать именно как результат отрицания нормальных унимодальных нечётких чисел.

В этом случае можно предложить подход, когда вся область определения функции принадлежности, у которой несколько локальных максимумов, разбивается на подмножества, в каждое из которых попадает один и только один локальный максимум. На этих подмножествах области определения производится дефаззификация обычным методом центра тяжести. В результате вычислений появляется набор чётких значений. Каждому из этих значений необходимо приписать некоторую степень уверенности в нём, так как в противном случае получается, что результатом дефаззификации стал вектор чётких значений, что несколько не соответствует определению процесса дефаззификации.

Самым простым способом расчёта степени уверенности в том или ином дефаззифицированном значении является деление площади под функцией принадлежности на подмножестве области определения на площадь под функцией принадлежности на всей области определения. Такие оценки для дефаззифицированных значений будут обладать хорошим качеством — нормированностью, то есть их сумма по всем оценкам будет равняться единице. Поэтому в некоторых задачах вместе с такими нормированными оценками можно использовать методы вероятностного подхода.

Другим способом, но менее адекватным, является получение значения функции принадлежности на дефаззифицированных значениях. Однако этот подход плох тем, что, во-первых, такие оценки не будут нормированными, а во-вторых, большинство из них будут иметь практически одинаковые значения, по крайней мере большинство из них будет иметь достаточно высокие значения уверенности.

3.1.1.4.2. Метод First-of-Maxima

В методе First-of-Maxima (первый максимум) чёткая величина выходной функции принадлежности находится как наименьшее значение, при котором достигается глобальный максимум дефаззифицируемой функции принадлежности.

$$z_0 = \min(z \mid C(z) = \max_u C(u)).$$

3.1.1.4.3. Метод среднего максимума

Этот метод является объединением двух предыдущих — из области максимизации конечной функции принадлежности выбирается центр тяжести, то есть средний максимум. Пусть G — максимизирующее подмножество дефаззифицируемой функции принадлежности, тогда для непрерывных вариантов средний максимум будет рассчитываться по формуле:

$$z_0 = \frac{\int_G z dz}{\int_G dz}.$$

То же для дискретных вариантов функций принадлежности:

$$z_0 = \frac{1}{N}\sum_{j=1}^{N} z_j.$$

Необходимо отметить, что этот метод даёт точку максимума для унимодальных нечётких чисел и среднюю точку максимума для толерантных нечётких чисел. Для функций принадлежности с несколькими локальными максимумами возникает такая же проблема, как и для метода центра тяжести (см. соответствующий раздел). Вполне возможно, что решение этой проблемы в рассматриваемом методе дефаззификации должно происходить подобно решению в методе центра тяжести.

3.1.1.4.4. Метод критерия максимума

Метод критерия максимума ненамного превосходит по адекватности метод первого максимума, так как в этом методе дефаззифицируемое значение получается при помощи случайного выбора из множества значений, максимизирующих функцию принадлежности:

$$z_0 \in \{z \mid C(z) = \max_u C(u)\}.$$

3.1.1.4.5. Метод высотной дефаззификации

Вероятно, что рассматриваемый метод является наиболее адекватным в условиях нечёткости, тем более что он в какой-то мере может быть подстроен под каждую конкретную задачу. Идея этого метода заключается в том, что в конечной функции принадлежности, которую необходимо дефаззифицировать, отсекаются все области, в которых функция принадлежности меньше некоего настраиваемого порога α. На оставшейся области (то есть на α-уровне дефаззифицируемой функции принадлежности) чёткое значение рассчитывается методом центра тяжести.

Если $\alpha = 0$, то этот метод перерождается в метод центра тяжести.

Этот метод подсказывает ещё один способ разрешения проблемы, сформулированной в описании метода центра тяжести. Если локальные максимумы имеют различные значения, то необходимо выбрать тот, у которого значение самое высокое. Затем выбрать уровень α, который равен максимуму, который стоит следующим в ряду убывающих значений максимума.

Если локальных максимумом в самой высокой оценкой принадлежности несколько, то решать эту проблему необходимо так же, как и в случае обычного метода центра тяжести.

3.1.1.5. Сопоставление нечётких переменных

В процессе вывода (как достоверного, так и нечёткого) одним из главнейших шагов является этап сопоставления. Все приведённые ранее примеры основаны на том, что во время сопоставления производится оценка равенства двух нечётких переменных, то есть определяется степень похожести функций принадлежности. Для пояснения механизма действия сопоставления в процессе нечёткого вывода этого было вполне достаточно. Однако при решении реальных задач на базах знаний, построенных с использованием знаний экспертов, оценка равенства двух нечётких переменных происходит в равной мере с оценкой их неравенства, а также различного рода сравнений («больше», «меньше или равно» и т. д.).

Например, в задачах планирования очень часто приходится оценивать продолжительность какой-либо работы при помощи нечёткого числа на шкале времени, при этом зачастую необходимо сравнивать с этим нечётким числом некоторые параметры, как обычные числовые, так и нечёткие [38].

В работах [38, 41] приводятся сводные классификации методов сравнения нечётких чисел, обобщая которые, можно получить следующую диаграмму:

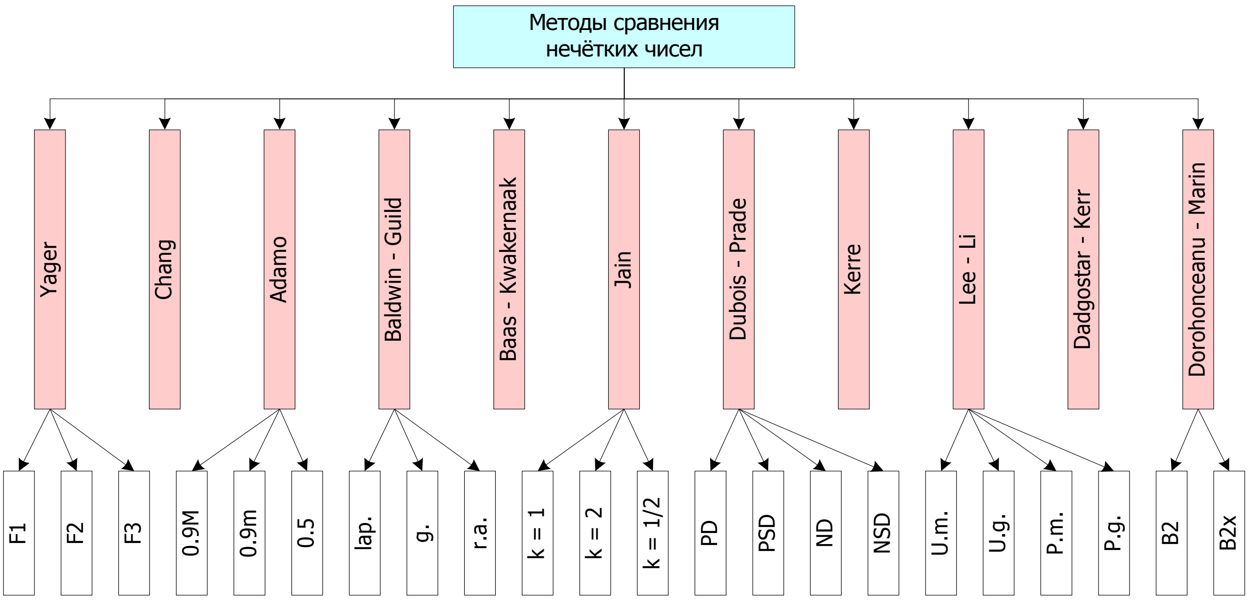

Рисунок 16. Обобщённая классификация методов сравнения нечётких чисел

На диаграмме представлены различные методы сравнения нечётких чисел, названные по именам авторов, разработавших эти методы (красные прямоугольники), а также варианты этих методов, если они существуют. Как видно из представленной диаграммы большинство методов сравнения имеют два и более вариантов, каждый из которых предназначен для тех или иных случаев сравнения. Однако в работе [41] показывается, что большинство из представленных методов обладают довольно серьёзными недостатками:

- большинство методов невосприимчивы к малым изменениям функций принадлежности;
- часто результаты, выдаваемые многими методами, не совпадают с ожиданиями интуиции человека;
- большинство методов заключаются в большом количестве вычислений, что пагубно сказывается на быстродействии программных средств;
- многие методы имеют несколько вариантов, которые зачастую выдают прямо противоположные результаты.

Менее всего таким недостаткам подвержены методы, разработанные в последнее время, где учтён предыдущий опыт. Такими методами являются метод частичных сравнений, который разработали Dadgostar и Kerr в 1996 году, а также простой метод сравнения, разработанный на основе пересмотра всех предыдущих методов Dorohonceanu и Marin.

Формализация задачи сравнения двух нечётких чисел заключается в том, что для нечёткого сравнения «A > B» необходимо вычислить степень его принадлежности нечёткому множеству «более, чем». Далее рассматривается часть методов сравнения нечётких чисел на основе отношения «больше», что, однако, не влияет на общность рассмотрения.

Остаётся отметить, что отношение «не равно» в своей сущности является отрицанием отношения «равно», которое вычисляется при помощи различного вида Т-норм. Для вычисления значения выражения «$A \neq B$», где A и B — некоторые нечёткие переменные, необходимо либо подвергнуть инверсии значение отношения «$A = B$», либо прибегнуть к помощи Т-конорм. Второй способ кажется более разумным [56].

3.1.1.5.1. Метод слабого сравнения

Для сравнения кусочно-линейных функций в работе [42] предлагается метод слабого сравнения, который основан на сравнении площадей областей под функциями принадлежности сравниваемых нечётких множеств. Этот метод, также как и большинство других методов сравнения нечётких чисел, не только даёт ответ на вопрос «Какое из чисел больше?», но и позволяет вычислять степень уверенности в полученном ответе.

Пусть заданы две функции принадлежности в кусочно-линейном представлении — A и B. Графики этих функций представлены на следующем рисунке:

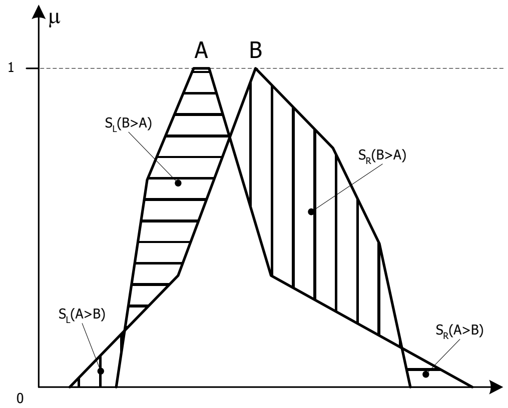

Рисунок 17. Сравнение нечётких чисел методом слабого сравнения

На представленном рисунке показаны области доминирования одной функции принадлежности над другой S_L и S_R, которые вычисляются по следующим формулам:

$$S_L(A>B)=\int\limits_{U(A,B)}(\inf_{x\in R}A_\alpha-\inf_{x\in R}B_\alpha)d\alpha$$
$$S_R(A>B)=\int\limits_{V(A,B)}(\sup_{x\in R}A_\alpha-\sup_{x\in R}B_\alpha)d\alpha,$$

где

$$U(A,V)=\{\alpha\,|\inf_{x\in R}A_\alpha\geq\inf_{x\in R}B_\alpha,\varepsilon\leq\alpha\leq1\}$$
$$V(A,V)=\{\alpha\,|\sup_{x\in R}A_\alpha\geq\sup_{x\in R}B_\alpha,\varepsilon\leq\alpha\leq1\}.$$

Значения S_L(B>A) и S_R(B>A) вычисляются аналогичным образом.

Таким образом, в соответствии с правилом слабого сравнения степень уверенности в том, что нечёткое число A больше нечёткого числа B, определяется следующим образом:

$$C(A>B)=\frac{\left(S_L(A>B)+S_R(A>B)\right)-\left(S_L(B>A)+S_R(B>A)\right)}{2}.$$

При помощи этого правила для нечётких чисел можно определить слабое неравенство, строгое неравенство и эквивалентность:

- $A\geq B\Leftrightarrow C(A>B)\geq0$;
- $A>B\Leftrightarrow C(A>B)>0$;
- $A\approx B\Leftrightarrow C(A>B)=C(B>A)$.

Остаётся отметить такое свойство полученных оценок уверенности относительно сравнимости нечётких чисел: C(A > B) = –C(B > A).

3.1.1.5.2. Простой метод сравнения нечётких чисел (Dorohonceanu и Marin)

Если области определения двух нормированных нечётких чисел A и B не пересекаются, то относительно этих двух чисел можно доподлинно сказать, какое из них меньше, а какое больше. Больше то, область определения которого лежит правее на оси действительных чисел **R**. Однако если области определения двух чисел пересекаются, или же не выполняется условие нормированности, то отношение «A больше, чем B» будет иметь некоторую степень достоверности в пределах (0; 1).

В последнем случае для вычисления степени достоверности выражения «A > B» можно воспользоваться формулой (метод B2):

$$\alpha_{A>B}=\frac{\sum\limits_{p=0}^{NOPL-1}\alpha_{A_p>B_p}\cdot length(A_p)\cdot length(B_p)}{\sum\limits_{p=1}^{NOPL-1}length(A_p)\cdot length(B_p)},$$

где:

- NOPL — количество уровней сечения соответствующих функций принадлежности;
- A_p и B_p — интервалы соответствия на уровне p для функций принадлежности чисел A и B;

- length (*interval*) — длина соответствующего интервала.

Степень достоверности отношения «A > B», вычисленная этим способом, обладает следующими важными свойствами:

$$
\begin{array}{l}
\alpha_{A>B} + \alpha_{B>A} = 1 \\
A \geq B \Leftrightarrow \alpha_{A>B} \geq \alpha_{B>A} \\
(A \geq B, B \geq C) \Rightarrow A \geq C, \alpha_{A>C} \geq \max(\alpha_{A>B}, \alpha_{B>C}) \\
(A \leq B, B \leq C) \Rightarrow \alpha_{A<C} \geq \max(\alpha_{A<B}, \alpha_{B<C})
\end{array}.
$$

Существует другой вариант этого метода, называемый B2x. В этом варианте одна из функций принадлежности разбивается на некоторое количество монотонных частей, то есть таких, где функция принадлежности либо только возрастает, либо только убывает, либо остаётся неизменной на всём интервале области определения. Для триангулярных чисел LR-типа таких интервалов монотонности будет два.

Далее все интервалы монотонности сравниваются со вторым нечётким числом. Результатом общего сравнения будет среднее арифметическое всех коэффициентов, полученных при сравнении частичных интервалов:

$$
\alpha_{A>B} = \frac{1}{n} \sum_{i=1}^{n} \alpha_{A_i>B}.
$$

Полученное таким образом значение также обладает всеми свойствами, перечисленными ранее.

3.1.1.5.3. Сравнение результатов методов

Следующая таблица обобщает результаты различных методов сравнения нечётких чисел на пяти тестовых примерах различной сложности. На рисунке, приведённом рядом с таблицей, показано графическое представление рассматриваемых тестовых примеров. На этом рисунке жирной сплошной линией показана функция принадлежности числа A_1, жирной пунктирной линией — A_2, точками — A_3 соответственно.

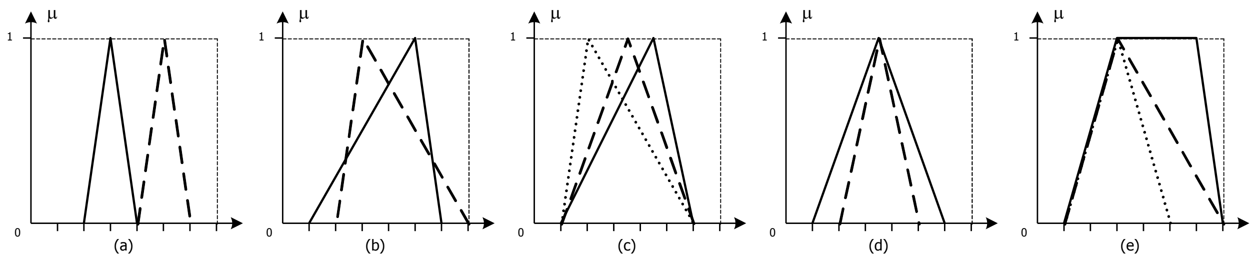

Рисунок 18. Графическое представление тестовых примеров для проверки методов сравнения нечётких чисел

Таблица 7. Результаты различных методов сравнения нечётких чисел

Метод	Вариант	(a)		(b)		(c)			(d)		(e)		
		A_1	A_2	A_1	A_2	A_1	A_2	A_3	A_1	A_2	A_1	A_2	A_3
Yager	F1	0.70	0.90	**0.61**	**0.53**	**0.76**	**0.70**	**0.63**	**0.50**	**0.50**	**0.62**	**0.56**	**0.50**
	F2	0.72	0.90	**0.66**	**0.69**	**0.90**	**0.76**	**0.66**	**0.61**	**0.54**	**0.81**	**0.64**	**0.58**
	F3	0.70	0.90	**0.58**	**0.56**	**0.80**	**0.70**	**0.60**	**0.50**	**0.50**	**0.62**	**0.54**	**0.50**
Chang		0.14	0.18	**0.40**	**0.34**	**0.46**	**0.41**	**0.38**	0.29	0.10	**0.56**	**0.33**	**0.20**
Adamo	0.9M	**0.71**	**0.91**	0.55	0.66	**0.91**	**0.73**	**0.55**	**0.53**	**0.51**	**0.81**	**0.54**	**0.52**
	0.9m	**0.71**	**0.91**	0.55	0.66	**0.91**	**0.73**	**0.55**	**0.53**	**0.51**	**0.81**	**0.54**	**0.52**
	0.5	**0.75**	**0.95**	0.75	0.72	**0.95**	**0.85**	**0.75**	**0.65**	**0.55**	**0.85**	**0.70**	**0.60**
Baldwin - Guild	lap.	**0.00**	**0.32**	0.42	0.33	**0.42**	**0.33**	**0.30**	**0.27**	**0.28**	**0.45**	**0.37**	**0.27**
	g.	**0.00**	**0.47**	0.44	0.37	**0.55**	**0.40**	**0.34**	**0.30**	**0.24**	**0.53**	**0.40**	**0.28**
	r.a.	**0.00**	**0.20**	0.34	0.24	**0.28**	**0.23**	**0.22**	**0.20**	**0.23**	**0.31**	**0.28**	**0.21**
Baas – Kwakernaak		**0.00**	**1.00**	**0.84**	**1.00**	**1.00**	**0.74**	**0.60**	**1.00**	**1.00**	1.00	1.00	1.00
Jain	k = 1	0.72	0.90	0.66	0.69	**0.90**	**0.76**	**0.66**	**0.73**	**0.67**	**0.90**	**0.69**	**0.64**
	k = 2	0.55	0.84	0.53	0.51	**0.84**	**0.65**	**0.54**	**0.60**	**0.48**	**0.82**	**0.56**	**0.45**
	k = ½	0.84	0.95	0.78	0.81	**0.85**	**0.86**	**0.78**	**0.83**	**0.80**	**0.94**	**0.80**	**0.77**
Dubois – Prade	PD	**0.00**	**1.00**	0.84	1.00	**1.00**	**0.74**	**0.60**	**1.00**	**1.00**	1.00	1.00	1.00
	PSD	**0.00**	**1.00**	0.54	0.46	**0.74**	**0.23**	**0.16**	0.73	0.24	0.80	0.20	0.00
	ND	**0.00**	**1.00**	0.54	0.46	**0.63**	**0.38**	**0.18**	0.27	0.76	0.50	0.50	0.50
	NSD	**0.00**	**1.00**	0.00	0.16	**0.26**	**0.00**	**0.00**	0.00	0.00	0.00	0.00	0.00
Kerre		0.80	1.00	**0.96**	**0.89**	**1.00**	**0.86**	**0.76**	**0.91**	**0.91**	**1.00**	**0.85**	**0.75**
Lee – Li	U.m.	**0.70**	**0.90**	0.61	0.53	**0.76**	**0.70**	**0.63**	0.50	0.50	**0.62**	**0.56**	**0.50**
	U.g.	—	—	—	—	—	—	—	0.12	0.04	—	—	—
	P.m.	**0.70**	**0.90**	0.53	0.58	**0.80**	**0.70**	**0.60**	0.50	0.50	**0.63**	**0.55**	**0.50**
	P.g.	—	—	—	—	—	—	—	0.09	0.03	—	—	—
Dadgostar – Kerr	PCM	**0.00**	**1.00**	**0.57**	**0.43**	**0.57**	**0.43**	**0.34**	**0.50**	**0.50**	**0.63**	**0.37**	**0.25**
Dorohonceanu – Marin	B2	**0.00**	**1.00**	**0.59**	**0.41**	**0.62**	**0.57**	**0.43**	**0.50**	**0.50**	**0.65**	**0.60**	**0.40**
	B2x	**0.00**	**1.00**	**0.59**	**0.41**	**0.64**	**0.59**	**0.41**	**0.50**	**0.50**	**0.68**	**0.61**	**0.39**

Каждое число в таблице показывает, с какой степенью соответствующее нечёткое число (показанное в заголовке столбца) больше остальных нечётких чисел в данном тестовом примере. Жирным шрифтом выделены интуитивно верные степени предпочтения одного нечёткого числа другом (другим).

Пример (a) показывает самый простой случай, когда области определения двух нечётких чисел не пересекаются. Интуиция подсказывает, что в этом случае число A_2 должно быть больше числа A_1 со степенью уверенности 1. Однако некоторые методы (Yager, Chang, Jain и Kerre) не дали чёткого результата в предпочтении одного числа другому.

Пример (b) представляет частичное пересечение двух нечётких чисел, осложнённое тем, что максимумы этих чисел лежат в необычных для такого рода пересечений местах. Некоторые варианты методов предпочли число A_1 (14 вариантов), другие — A_2 (9 вариантов). Небольшое число методов (например, Adamo 0.9M) не смогли вычислить чёткое предпочтение. Такая ситуация вносит больше неясности, чем помогает разрешить задачу сравнения.

Пример (c) оказался простым для всех методов, каждый из которых правильно расположил предпочтение нечётким числам: $A_1 > A_2 > A_3$.

Пример (d) представляет случай, когда оба нечётких числа достигают максимума в одной и той же точке, однако имеют разную дисперсию. Некоторые методы предпочли число A_1, но остальные не смогли различить оба нечётких числа. В реальных задачах такой случай должен рассматриваться с точки зрения проблемной области, то есть разрешение такого конфликта должно быть проблемно-зависимым.

В примере (e) большинство методов предпочли число A_1, но некоторые из них не смогли различить эти числа (Baas — Kwakernaak и два варианта метода Dubois — Prade), хотя интуиция подсказывает, что число A_1 всё-таки больше остальных.

Таким образом, наиболее верные результаты показали методы Dadgostar — Kerr и Dorohonceanu — Marin, которые, кроме того, лишены некоторых недостатков остальных методов, так как основаны на наиболее поздних разработках аппарата нечёткой логики [40]. В частности, оба этих метода являются чрезвычайно чувствительными к малым изменениям функций принадлежности сравниваемых нечётких чисел.

3.1.1.5.4. Сравнение кусочно-линейных функций принадлежности

Основная идея при сравнении кусочно-линейных функций принадлежности (равно как и функций принадлежности вообще) заключается в использовании механизма дефаззификации при сравнении. Самый простой метод заключается в сравнении дефаззифицированных значений, вместо сравнения самих функций принадлежности. Однако основной проблемой в данном случае является необходимость подбора адекватного метода дефаззификации, который применим в конкретном случае.

Другой важной идеей в сравнении кусочно-линейных функций принадлежности является сравнение одной функции с простым числом из области определения рассматриваемой функции. Эта идея является расширением определения обычного арифметического сравнения — результат вычитания одного числа из другого сравнивается с нулём. То же самое можно сделать и в нечёткой арифметике. При заданной операции вычитания можно найти разность сравниваемых функций принадлежности и сравнить эту разность с нулём.

При сравнении функции принадлежности нечёткого множества с точкой на числовой оси, на которой определена сравниваемая функция принадлежности, необходимо произвести дефаззификацию методом центра тяжести, так как этот метод даёт наиболее адекватную оценку. Полученное чёткое число после дефаззификации уже можно сравнивать с исходным методами обычной арифметики.

Иногда в процессе решения задач бывает необходимо получить степень уверенности в том, что одно нечёткое число больше (меньше) второго. Такую степень уверенности можно получить, сравнивая площади под функцией принадлежности, которые находятся слева и справа от числа, с которым происходит сравнение. Например, на следующем рисунке графически показан процесс сравнения некоторого нечёткого числа с нулём:

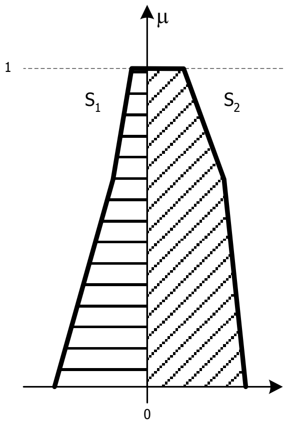

Рисунок 19. Сравнение функции принадлежности с нулём

На этом рисунке площади под функцией принадлежности слева и справа от нулевой ординаты обозначены как S_1 и S_2 соответственно. После вычисления этих площадей (которое в случае кусочно-линейного представления функций принадлежности заключается в простом численном интегрировании методом трапеций), степени уверенности в том, что представленное нечёткое число больше или меньше нуля равны:

$$\begin{aligned}\mu(F \leq 0) &= \frac{S_1}{S_1 + S_2} \\ \mu(F \geq 0) &= \frac{S_2}{S_1 + S_2}\end{aligned}.$$

Остаётся отметить, что полученные таким образом степени уверенности являются нормированными — их сумма равна единице, что в некоторых задачах позволяет интерпретировать этот результат с вероятностной точки зрения.

3.1.1.6. Нечёткая арифметика

Нечёткая арифметика определяет арифметические операции над нечёткими числами в частности и над нечёткими множествами (их функциями принадлежности) в общем. Такие нечёткие арифметические операции довольно часто проявляются при решении реальных задач, куда проникла нечёткость:

1. Чему равно расстояние, пройдённое автомобилем на «большой» скорости за полчаса езды? Чему равно расстояние, пройдённое автомобилем на «обычной» скорости за «небольшое» время?

2. Какова площадь помещения, длина которого равна примерно пяти метрам, а ширина не превышает 6 метров?

3. Какова длина стержня, сваренного из двух малых стержней, длина первого из которых находится около десяти сантиметров, а длина второго чуть больше пятнадцати сантиметров?

Любая из четырёх основных арифметических операций над нечёткими множествами может быть описана при помощи принципа расширения Заде [56]:

$$\mu_{A*B}(z) = \max_{x*y=z}\{\min[\mu_A(x), \mu_B(y)]\},$$

где A и B — два нечётких множества, определённых на X и Y соответственно. При этом вполне естественно требование, что на элементах множеств X и Y должны быть определены арифметические операции $* \in \{+,-,\times,/\}$.

В случае если в процессе вывода используется другая операция нечёткой конъюнкции, а не взятие минимума, как показано в предыдущей формуле, то вместо операции min в формулу можно подставить любую операцию из класса Т-норм [48]. Естественно, что использование такой операции должно быть обосновано и адекватно представлениям с точки зрения здравого смысла.

3.1.1.6.1. Арифметика над шеститочечными нечёткими числами

Так как шеститочечное представление нечётких чисел включает в себя в качестве подмножества триангулярные и трапециевидные нечёткие числа, то рассмотрение нечётких арифметических операций без потери общности можно производить именно на шеститочечных нечётких числах. В этом случае формула для вычисления значения арифметической операции $* \in \{+,-,\times,/\}$ принимает вид:

$$\forall \alpha \in (0;1]: C_\alpha = (A * B)_\alpha = A_\alpha * B_\alpha,$$

где X_α — α-срез нечёткого множества X (см. Глоссарий).

Эта формула показывает, что α-срез результата арифметической операции над двумя нечёткими множествами равен арифметической операции над α-срезами этих нечётких множеств. Принимая во внимание определение шеститочечного нечёткого числа, можно записать [42]:

$$\begin{aligned}
A + B &= \left\{\underline{a}^{\varepsilon} + \underline{b}^{\varepsilon}, \underline{a}^{\lambda} + \underline{b}^{\lambda}, \underline{a} + \underline{b}, \overline{a} + \overline{b}, \overline{a}^{\lambda} + \overline{b}^{\lambda}, \overline{a}^{\varepsilon} + \overline{b}^{\varepsilon}\right\} \\
A - B &= \left\{\underline{a}^{\varepsilon} - \underline{b}^{\varepsilon}, \underline{a}^{\lambda} - \underline{b}^{\lambda}, \underline{a} - \underline{b}, \overline{a} - \overline{b}, \overline{a}^{\lambda} - \overline{b}^{\lambda}, \overline{a}^{\varepsilon} - \overline{b}^{\varepsilon}\right\} \\
A \times B &= \left\{\underline{a}^{\varepsilon} \times \underline{b}^{\varepsilon}, \underline{a}^{\lambda} \times \underline{b}^{\lambda}, \underline{a} \times \underline{b}, \overline{a} \times \overline{b}, \overline{a}^{\lambda} \times \overline{b}^{\lambda}, \overline{a}^{\varepsilon} \times \overline{b}^{\varepsilon}\right\} \\
A / B &= \left\{\underline{a}^{\varepsilon} / \underline{b}^{\varepsilon}, \underline{a}^{\lambda} / \underline{b}^{\lambda}, \underline{a} / \underline{b}, \overline{a} / \overline{b}, \overline{a}^{\lambda} / \overline{b}^{\lambda}, \overline{a}^{\varepsilon} / \overline{b}^{\varepsilon}\right\}
\end{aligned}.$$

Операция взятия минимума и максимума на шеститочечных нечётких числах определяется абсолютно точно также.

3.1.1.6.2. Арифметика над кусочно-линейными функциями принадлежности

Как известно, кусочно-линейные функции принадлежности определяются как непустое множество упорядоченных пар вида $\langle x_i, \mu(x_i)\rangle_{i=\overline{1,n}}$. Над таким представлением функций принадлежности очень легко производить нечёткие арифметические операции, так как для каждой операции необходимо получать значения только в заданных точках, а промежуточные значения находятся при помощи линейной интерполяции.

В соответствии с принципом расширения Заде [56] нечёткая арифметическая операция производится над декартовым произведением множества точек двух кусочно-линейных функций принадлежности [48]. Пусть имеются две функции принадлежности:

$$MF_1 = \left\{\left\langle x_i^1, \mu_1(x_i^1)\right\rangle\right\}_{i=1}^{m}$$
$$MF_2 = \left\{\left\langle x_j^2, \mu_2(x_j^2)\right\rangle\right\}_{j=1}^{n}.$$

Для того чтобы найти значение какой-либо нечёткой арифметической операции над этими функциями принадлежности, необходимо построить матрицу размера m × n, каждый элемент которой будет равен значению арифметической операции на величинах $x_i^1, i = \overline{1,m}$ и $x_j^2, j = \overline{1,n}$. Каждому полученному значению приписывается следующая степень принадлежности:

$$\mu^*(x_i^1 * x_j^2) = \min\left(\mu_1(x_i^1), \mu_2(x_j^2)\right).$$

Среди элементов построенной матрицы могут встретиться повторяющиеся значения. Для всех таких наборов из повторяющихся значений необходимо выбрать максимальную величину степени принадлежности μ*. Эта степень принадлежности и будет являться окончательной для рассматриваемого значения арифметической операции.

Этот процесс проще всего рассматривать на конкретном примере. Пусть необходимо вычислить суммарное время ожидания, составленное из величин с названиями терм-множеств «небольшое» и «около часа». Пусть эти терм-множества описываются следующими функциями принадлежности:

$$\textit{небольшое} = \{(0;1.0), (10;0.95), (15;0.9), (20,0.75), (30;0.2), (45;0)\}$$
$$\textit{около часа} = \{(30;0), (40;0.25), (50;0.9), (60;1), (70;0.9), (80;0.25), (90;0)\},$$

где время показано в минутах. Графики функций принадлежности этих нечётких множеств показаны на следующем рисунке:

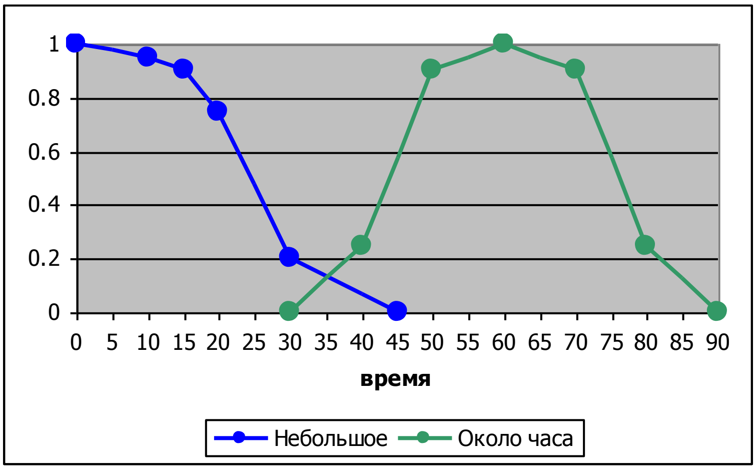

Рисунок 20. Графики функций принадлежности терм-множеств «небольшое» и «около часа»

Для того чтобы получить сумму этих двух нечётких множеств, необходимо составить таблицу (матрицу) размера 6 × 7, в ячейках которой будут записаны суммы соответствующих точек на оси времени и приписанные к ним степени принадлежности. Эта матрица показана в следующей таблице:

Таблица 8. Значения суммы точек двух функций принадлежности на их декартовом произведении с приписанными степенями принадлежности

	0	**10**	**15**	**20**	**30**	**45**
30	30 / 0.00	40 / 0.00	45 / 0.00	50 / 0.00	60 / 0.00	75 / 0.00
40	40 / 0.25	50 / 0.25	55 / 0.25	60 / 0.25	70 / 0.20	85 / 0.00
50	50 / 0.90	60 / 0.90	65 / 0.90	70 / 0.75	**80 / 0.20**	95 / 0.00
60	60 / 1.00	70 / 0.95	75 / 0.90	**80 / 0.75**	90 / 0.20	105 / 0.00
70	70 / 0.90	**80 / 0.90**	85 / 0.90	90 / 0.75	100 / 0.20	115 / 0.00
80	**80 / 0.25**	90 / 0.25	95 / 0.25	100 / 0.25	110 / 0.20	125 / 0.00
90	90 / 0.00	100 / 0.00	105 / 0.00	110 / 0.00	120 / 0.00	135 / 0.00

Далее в соответствии с алгоритмом в таблице необходимо выделить одинаковые значения суммы на парах точек. Для примера взято значение «80 минут», выделенное в таблице отличающимся фоном и полужирным шрифтом. Из четырёх степеней принадлежности, приписанных значению «80 минут» необходимо выбрать одно. Принцип расширения Заде утверждает, что это должно быть максимальная степень принадлежности, то есть в рассматриваемом случае — 0.90. Ту же самую операцию необходимо проделать для каждого полученного значения, после чего выводится новая функция принадлежности, равная сумме терм-множеств «небольшое» и «около часа»:

$$\Sigma = \left\{ \begin{array}{l} (30;0),(40;0.25),(45;0),(50;0.9),(55;0.25),(60;1),(65;0.9),(70;0.95), \\ (75;0.9),(80;0.9),(85;0.9),(90;0.75),(95;0.25),(100;0.25),(105;0), \\ (110;0.2),(115;0),(120;0),(125;0),(130;0),(135;0) \end{array} \right\}$$

График полученной суммы двух функций принадлежности показан на следующем рисунке:

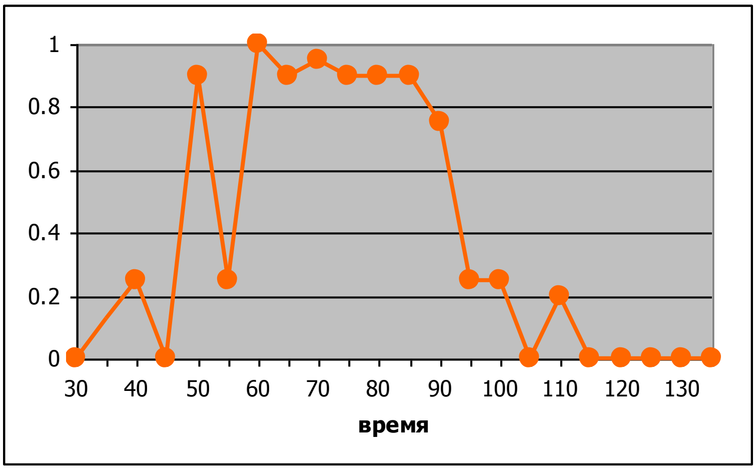

Рисунок 21. График функции принадлежности суммы двух нечётких множеств

Как видно на графике, полученная сумма не совсем адекватно отражает то, что можно было бы ожидать в соответствии с интуицией. Особенно это касается точек «45 минут», «55 минут» и «105 минут». То же самое можно сказать и о точке «65 минут». Значения функции принадлежности суммы во всех упомянутых точках выбиваются из общего ряда и нарушают монотонность функции. Здравый смысл подсказывает, что значения суммы в этих точках необходимо либо линейно интерполировать, либо вычислять каким-либо другим способом, а не в соответствии с принципом расширения Заде. Наиболее простой способ заключается в линейной интерполяции значений функции принадлежности суммы в этих точках, то есть простое удаление записей о них в представлении результата суммирования (так как промежуточные значения в кусочно-линейном представлении вычисляются при помощи линейной интерполяции).

Критерием поиска таких «нехороших» точек может быть различие знаков производных слева и справа от точки. Если знаки производных различны, то есть в точке нарушается монотонность функции, то следует задуматься о том, что точка может быть «нехорошей». Однако этот критерий выявляет не только такие точки, но и точки экстремумов, которые обязательно должны входить в результирующую функцию принадлежности. Например, точка «60 минут» не может быть удалена из представления суммы двух нечётких множеств, так как в этой точке сумма достигает своего максимума, хотя предложенный критерий показывает эту точку, как «нехорошую».

Другой проблемой при проведении арифметических операций над нечёткими множествами является появление областей ступенчатости результирующих функций принадлежности. Так, например, в рассматриваемом примере, наиболее характерной областью ступенчатости является интервал [«115 минут»; «135 минут»], на всём протяжении которого функция принадлежности принимает значение 0. Другими интервалами ступенчатости являются следующие: [«75 минут»; «85 минут»] и [«95 минут»; «100 минут»]. Это не такая серьёзная проблема, как нарушение монотонности функции принадлежности, однако использование более плавных функций более адекватно отражает реальные значения нечётких параметров [23, 42, 48].

Решение этой проблемы также заключается в отбрасывании из представления результата лишних точек. В каждом интервале ступенчатости необходимо оставить только одну точку, остальные должны вычислять на общих основаниях с использованием линейной интерполяции. Если интервал ступенчатости находится в области возрастания функции принадлежности, то необходимо оставить нижнюю границу интервала. И обратно — если интервал ступенчатости находится в области убывания функции, то необходимо оставить верхнюю границу интервала.

Необходимо отметить, что в приведённом примере нельзя просто удалить интервал ступенчатости [«115 минут»; «135 минут»], где функция принадлежности имеет значение 0, так как в соответствии с арифметикой интервалов (см. далее) областью определения функции принадлежности суммы является сумма областей определения слагаемых, то есть в рассматриваемом примере:

$$[0;45]+[30;90]=[30;135].$$

Проведя все удаления лишних точек, можно получить следующую функцию принадлежности:

$$\Sigma^* = \begin{Bmatrix} (30;0),(40;0.25),(50;0.9),(60;1),(70;0.95), \\ (85;0.9),(90;0.75),(100;0.25),(110;0.2),(135;0) \end{Bmatrix}$$

График этой функции принадлежности приведён на следующем рисунке:

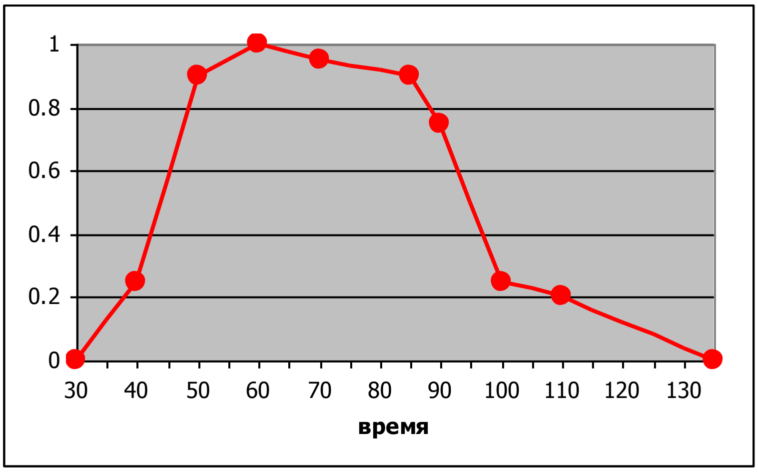

Рисунок 22. График преобразованной функции принадлежности суммы двух нечётких множеств

Как видно из представленного рисунка, полученная в результате удаления «нехороших» точек и интервалов ступенчатости функция более гладкая (если этот термин можно применять к кусочно-линейным функциям).

Таким образом, предложенная арифметика над кусочно-линейными функциями может применяться не только к нечётким числам, но и в общем случае к функциям принадлежности, построенным на любых числовых осях, то есть таких осях, к элементам которых могут быть применены арифметические операции.

3.1.1.7. Нечёткая арифметика и здравый смысл

Иногда нечёткая арифметика, описанная в предыдущем разделе, даёт не те результаты, которые ожидает здравый смысл [48]. Например, пусть некоторый человек весит около сотни килограммов. Сколько будет весить этот человек, если наберёт в весе ещё один килограмм? Интуиция подсказывает, что нечёткое выражение его веса не должно измениться — он всё также будет весить «около сотни килограммов». Однако правила нечёткой арифметики дадут другой результат.

Ещё более обескураживающий пример. Динозавр, живший около четырнадцати миллионов лет назад, через пять лет всё также должен быть динозавром, который жил около четырнадцати миллионов лет назад. В общем, рассматривая такие примеры, можно прийти к выводу, что если число a намного больше числа b (то есть $a >> b$), то при сложении нечёткого числа «примерно a» с числом b в результате всё также должно быть нечёткое число «примерно a».

В работах [48, 44] предложен формализм для осуществления таких операций над нечёткими числами. Предлагается следующий подход:

- Вычислить сумму двух нечётких чисел: $C = A + B$.
- Определить интервал возможных значений на области определения множества C (например, все такие x, для которых значение функции принадлежности $\mu_C(x)$ больше какого-либо порога α).
- Выбрать «самое типичное» значение на этом интервале — c.
- Результатом вычисления суммы A и B будет терм-множество «примерно c».

Однако главной проблемой такого алгоритма является выбор «самого типичного» значения. Интуитивно ясно, что таким значением будет самое «простое» число в рассматриваемом интервале (простое не в традиционном математическом смысле, а в смысле восприятия его человеком — например, круглые числа просты для восприятия и запоминания людьми). Но как формально определить самое «простое» число?

Предлагается воспользоваться модифицированным понятием «сложность Колмогорова», которое определяется как минимальный размер формулы, необходимой для определения заданного числа (в то время, как исходная сложность Колмогорова определяется, как минимальный размер программы, необходимый для вычисления заданного числа). Так как в вычислении чисел необходимости нет, а есть необходимость только в оценке его сложности, вводится именно модифицированная сложность Колмогорова [48].

Вновь введённые понятия усложняют процесс вычисления результатов нечётких операций (по крайней мере, для сложения и вычитания). Кроме того, сложение перестаёт быть ассоциативным. И самая главная проблема заключается в том, что ещё не предложено алгоритма, который сможет найти «самое типичное» число в заданном интервале.

3.1.2. Вывод в условиях неопределённости

Для вывода в условиях неопределённости разработано довольно большое множество различных методов и формализмов, начиная от различного рода вероятностных подходов [14], до специализированных теорий. Из последних наиболее широкое применение получили теория Байеса [11], её расширение — схема Пиэрла, и теория Демпстера-Шейфера [39, 52].

Для оперирования неопределённостью в продукциях необходимо задать функции пересчёта, позволяющие вычислять:

- Меру неопределённости посылки правила по мерам неопределённости составляющих его высказываний (фактов).
- Меру неопределённости заключения правила по мере неопределённости самого правила и мере неопределённости его посылки.
- Объединённую меру неопределённости высказывания по мерам неопределённости правил, заключением которых является рассматриваемое высказывание.

3.1.2.1. Метод Байеса

Метод Байеса заключается в использовании эвристического расширения формулы Байеса, взятой из классической теории вероятности:

$$\frac{P(H / E_1 \& E_2)}{P(\sim H / E_1 \& E_2)} = \frac{P(H)}{P(\sim H)} * \frac{P(E_1 / H)}{P(E_1 / \sim H)} * \frac{P(E_2 / H)}{P(E_2 / \sim H)}.$$

Корректное применение этой формулы возможно только при одновременной независимости E_1 и E_2 по H и ~H, что является очень сильным требованием.

Метод Байеса нашёл своё развитие в расширенных сетях Байеса, схеме Пиэрла и методе Демпстера-Шейфера, которые дают более впечатлительные результаты, однако являются более тяжёлыми для обработки.

3.1.2.2. Обработка неопределённости методом Демпстера-Шейфера

Теория Демпстера-Шейфера была разработана с целью обобщения вероятностного подхода к описанию неопределённости и связана с попыткой освободиться от аксиом теории вероятности при описании субъективной веры эксперта [52].

Важнейшим элементом теории Демпстера-Шейфера является операция объединения свидетельств, задаваемая правилом Демпстера [39]. Правило является эвристическим и основывается на интуитивных соображениях о разумном способе объединения не связанных друг с другом свидетельств. Недостатком метода является вычислительная сложность правила объединения свидетельств, которая экспоненциально зависит от мощности множества объединяемых свидетельств.

В теории Демпстера-Шейфера рассматриваются интервальные оценки неопределённости вида [n, p] для факта A, где:

- $n = Bel_A(v)$ — уверенность в том, что A = v.
- $p = Pl_A(v)$ — возможность того, что A = v.

Первое правило, необходимое для расчёта коэффициентов неопределённости сложного высказывания по составляющим его простым, представлено следующей таблицей.

Пусть имеются два утверждения: E_1 и E_2, имеющие коэффициенты неопределённости $[n_2, p_2]$ и $[n_1, p_1]$ соответственно.

Таблица 9. Вычисление сложных мер неопределённости в теории Демпстера-Шейфера

Подход \ Оператор	$\sim E_1$	$E_1 \vee E_2$	$E_1 \wedge E_2$
E_1 и E_2 независимы	$1 - p_1$ $1 - n_1$	$n_1 + n_2 - n_1 * n_2$ $p_1 + p_2 - p_1 * p_2$	$n_1 * n_2$ $p_1 * p_2$
Нет информации о зависимости E_1 и E_2	$1 - p_1$ $1 - n_1$	$\max [n_1, n_2]$ $\min [1, p_1 + p_2]$	$\max [0, n_1 + n_2 - 1]$ $\min [1, p_1 + p_2]$
E_1 и E_2 взаимно исключают друг друга	$1 - p_1$ $1 - n_1$	$n_1 + n_2$ $p_1 + p_2$	0 0
$E_1 \rightarrow E_2$ или $E_2 \rightarrow E_1$	$1 - p_1$ $1 - n_1$	$\max [n_1, n_2]$ $\max [p_1, p_2]$	$\min [n_1, n_2]$ $\min [p_1, p_2]$

Если утверждение правила сложное, то есть состоит из отрицания, конъюнкции или дизъюнкции нескольких утверждений, то по этой таблице высчитываются общие коэффициенты неопределённости посылки, которые затем используются для вычисления коэффициентов неопределённости заключения правила.

Правило типа $E \rightarrow H$ определяется двумя оценками:

- $[Bel (E \rightarrow H), Pl (E \rightarrow H)]$
- $[Bel (\sim E \rightarrow H), Pl (\sim E \rightarrow H)]$.

Пусть E имеет коэффициенты [s, t], а продукция $E \rightarrow H$ имеет коэффициенты $[n_1, p_1]$ и $[n_2, p_2]$. Тогда коэффициенты неопределённости для H вычисляются по формулам:

$$P(H) \geq \begin{cases} n_1 \cdot s + n_2 \cdot (1-s), \text{ если } n_1 \geq n_2 \\ n_1 \cdot t + n_2 \cdot (1-t), \text{ иначе} \end{cases},$$

$$P(H) \leq \begin{cases} p_1 \cdot s + p_2 \cdot (1-s), \text{ если } p_1 \leq p_2 \\ p_1 \cdot t + p_2 \cdot (1-t), \text{ иначе} \end{cases}.$$

Эти формулы представляют собой метод вычисления коэффициентов неопределённости цели правила по коэффициентам неопределённости посылки правила и самого правила — так называемое «правило Демпстера».

3.1.2.2.1. Объединение свидетельств

Объединение свидетельств производится в тех случаях, когда для одного и того же факта были вычислены значения по разным продукциям с соответственно разными значениями коэффициентов неопределённости.

Пусть первое правило дало для H оценку $[n_1, p_1]$, а второе $[n_2, p_2]$. Тогда общая оценка коэффициентов неопределённости, помещаемая в рабочую память машины вывода, имеет вид:

$$n = \frac{n_1 p_2 + n_2 p_1 - n_1 n_2}{1 - n_1(1-p_2) - n_2(1-p_1)},$$

$$p = \frac{p_1 p_2}{1 - n_1(1-p_2) - n_2(1-p_1)}.$$

Приведённые выше формулы являются общими, однако в процессе обработки неопределённых знаний можно пользоваться эвристическими расширениями этих формул, построенными на основе частных видов правил. Это позволяет ускорить процесс обработки коэффициентов неопределённости методом Демпстера-Шейфера.

Необходимо отметить, что отношение, используемое в посылке правила, не имеет значения при расчёте коэффициентов неопределённости, так как теория Демпстера-Шейфера основана на том, что эти коэффициенты в общем случае не зависят от вида продукции. В заключениях продукций находятся действия, выполнение которых ведёт к означиванию переменных, хранимых в рабочей памяти решателя.

3.1.2.2.2. Различные виды правил

Правила общего вида

ЕСЛИ X = V_1 [s, t] (E)
  ТО Y = V_2 $[n_1, p_1]$ (H_1)
  ИНАЧЕ Z = V_3 $[n_2, p_2]$ (H_2)

Это самый общий вид правил. Требуется рассчитать коэффициенты, помещаемые в рабочую память для фактов Y = V_2 и Z = V_3.

- Bel (E $\rightarrow$ H_1) = n_1, Pl (E $\rightarrow$ H_1) = p_1.
- Bel (~E $\rightarrow$ H_1) = 0, Pl (~E $\rightarrow$ H_1) = 1.

Соответственно по правилу Демпстера можно получить оценку неопределённости для P (H_1):

$$P(H_1) \in [n_1 s, 1 - s + p_1 s].$$

Расчёт коэффициентов для H_2 аналогичен, принимая во внимание, что данное правило можно переписать в виде:

ЕСЛИ X != V_1 [s, t] (~E)
ТО Z = V_3 [n_2, p_2] (H_2)

И тогда:

$$P(H_2) \in [n_2 - n_2 t, t + p_2 - p_2 t].$$

Правило с одинаковыми фактами в частях ТО и ИНАЧЕ

ЕСЛИ X = V_1 [s, t] (E)
ТО Y = V_2 [n_1, p_1] (H_1)
ИНАЧЕ Y = V_3 [n_2, p_2] (H_2)

В этом случае напрямую применяется правило Демпстера для расчётов коэффициентов неопределённости, так как:

- Bel (E $\rightarrow$ H_1) = n_1, Pl (E $\rightarrow$ H_1) = p_1.
- Bel (~E $\rightarrow$ H_1) = n_2, Pl (~E $\rightarrow$ H_1) = p_2.

Два применимых правила с одинаковыми консеквентами

ЕСЛИ X = V_1 [s_1, t_1] (E_1)
ТО Y = V_2 [n_1, p_1] (H)

ЕСЛИ Z = V_3 [s_2, t_2] (E_2)
ТО Y = V_2 [n_2, p_2] (H)

Эти две продукции представляют собой случай объединения свидетельств, то есть если оба правила выполняются, то в рабочей памяти решателя должна быть проведена корректировка коэффициентов неопределённости для факта H. К данному случаю машина вывода будет приходить тогда, когда в заключении правила есть факт H, который уже имеется в рабочей памяти вместе с рассчитанными ранее коэффициентами неопределённости.

По формулам объединения свидетельств имеет место:

$$P(H) \geq \frac{(n_1 s_1)(1 + p_2 s_2 - s_2) + (n_2 s_2)(1 + p_1 s_1 - s_1) - (n_1 s_1)(n_2 s_2)}{1 - (n_1 s_1)(s_2 - p_2 s_2) - (n_2 s_2)(s_1 - p_1 s_1)},$$

$$P(H) \leq \frac{(1 + p_1 s_1 - s_1)(1 + p_2 s_2 - s_2)}{1 - (n_1 s_1)(s_2 - p_2 s_2) - (n_2 s_2)(s_1 - p_1 s_1)}.$$

С помощью приведённых формул можно рассчитать коэффициенты неопределённости для любых правил и их последовательностей. Алгоритм расчёта следующий:

1. Рассчитать коэффициенты неопределённости посылки правила. Если посылка правила сложная, то есть состоящая из нескольких простых высказываний, связанных операциями отрицания, конъюнкции и дизъюнкции, то расчёт общих коэффициентов неопределённости производится по таблице вычисления сложных мер.

2. Рассчитать коэффициенты для заключения правила, используя подсчитанные значения на предыдущем шаге и коэффициенты неопределённости самого правила. Данный расчёт производится по формулам для первого и второго типа правил.

3. Если в рабочей памяти решателя уже находится факт, для которого рассчитывались коэффициенты на шаге 2, то необходимо произвести объединение свидетельств, производимое по формулам для третьего вида правил.

3.1.3. Вывод на неточных и недоопределённых знаниях

Как уже неоднократно утверждалось, некоторые виды неточности и недоопределённости могут быть представлены в виде интервалов на действительной оси **R**. В подавляющем большинстве случаев такое представление будет иметь вид непрерывного интервала, то есть, по сути, состоять из двух действительных чисел — нижней и верхней границ интервала (см. раздел о представлении неточности и недоопределённости). Поэтому для обработки знаний с рассматриваемыми НЕ-факторами необходимо использовать интервальную арифметику, в рамках которой происходит сравнение интервалов и вычисление над ними арифметических операций.

Как показано в работе [3] наиболее перспективными для нахождения решений в системах с неточностью являются интервальные [15, 47] методы. Эти методы получили большое распространение при решении систем дифференциальных уравнений, систем линейных и нелинейных уравнений, задач глобальной оптимизации [51].

Применение интервального анализа и различных минимаксных (гарантированных) подходов обладает целым рядом преимуществ [15]:

- Не требуется знание вероятностных характеристик неопределённых факторов, которые редко бывают точно известны на практике.
- При минимаксном подходе получаются строгие оценки для самих искомых величин, а не для вероятностей или математических ожиданий, что имеет большое значение при наличии малого числа замеров параметров и одной или нескольких реализаций.
- Статистические характеристики не могут гарантировать определённый исход одного конкретного опыта.
- Во всех случаях даются гарантированные двусторонние аппроксимации искомых решений.

В общем случае точность интервального результата полностью определяется следующими четырьмя факторами [2]:

- Неопределённостью в задании исходных данных.
- Округлениями при выполнении операций, изменяющих или порождающих интервальные объекты.
- Приближённым характером используемого численного метода.
- Степенью учёта зависимостей между участвующими в вычислении интервальными объектами (переменными и константами).

Увеличение точности расчётов (уменьшение ширины результирующего интервала) достигается за счёт компенсации влияния этих факторов. Задача получения для данного множества машинно-представимых чисел самого узкого интервала, содержащего объединённое расширение соответствующей рациональной функции, может ставиться как оптимизационная [3].

Для уменьшения погрешности округления используются изменение разрядности чисел, различные способы машинного представления и специальное упорядочение цепочки следующих друг за другом операций. А компенсация влияния четвёртого фактора осуществляется путём предварительной обработки исходного алгоритма, в процессе вычислений и апостериорно [2].

Далее последовательно рассматриваются вопросы сравнения интервальных величин, а также некоторые математические формализмы и аппараты для вычисления арифметических операций над интервалами.

3.1.3.1. Интервальная арифметика

Для того чтобы начать рассмотрение аппарата интервальной арифметики, необходимо формализовать некоторые понятия, отношения между ними и операции над этими понятиями. Главным понятием интервальной математики вообще является понятие интервала, а также интервального числа [3].

3.1.3.1.1. Интервальные числа

Пусть **R** — множество всех действительных чисел. Под интервалом [a; b], a ≤ b понимается замкнутое ограниченное подмножество **R** вида:

$$[a;b] = \{x \mid x \in R, a \leq x \leq b\}.$$

Множество всех интервалом обозначается в виде I(**R**). Элементы множества I(**R**) будут обозначаться заглавными буквами, а нижняя и верхняя граница интервала будут обозначаться соответствующими строчными буквами:

$$A \in I(R), A = [a, \overline{a}].$$

Таким образом, элементы множества I(**R**) называются интервальными числами. Соответственно само множество I(**R**) можно рассматривать как расширение над множеством действительных чисел.

Символы ($\in$), ($\cap$), ($\subset$) и т. п. понимаются в обычном теоретико-множественном смысле, причём символ ($\subset$) обозначает не только строгое включение, но и равенство интервалов. Два интервала A и B равны тогда и только тогда, когда равны их нижние и верхние границы соответственно $(a = b, \overline{a} = \overline{b})$.

Отношение порядка на множестве I(**R**) определяется следующим образом: A < B тогда и только тогда, когда $\overline{a} < b$. Возможно так же упорядочение по включению: A не превосходит B, если A $\subset$ B. Обычно используется первое определение.

Пересечение A $\cap$ B интервалов A и B пусто, если A < B или B < A, в противном случае $A \cap B = [\max(a, b), \min(\overline{a}, \overline{b})]$ — интервал из множества I(**R**).

Симметричным, по определению, является интервал $[a, \overline{a}]$, для которого $-a = \overline{a}$.

Шириной ω(A) интервала A называется величина $\omega(A) = \overline{a} - a$.

Середина m(A) есть полусумма концов интервала A: $m(A) = \frac{a + \overline{a}}{2}$.

Абсолютная величина |A| определяется как $|A| = \max(|a|, |\overline{a}|)$.

Расстояние ρ(A, B) между элементами A, B ∈ I(**R**) определяется равенством:

$$\rho(A, B) = \max(|a - b|, |\overline{a} - \overline{b}|).$$

Вырожденный интервал, то есть интервал с совпадающими концами $a = \overline{a}$, отождествим с самим действительным числом a. Таким образом, как уже было сказано — **R** ∈ I(**R**).

Интервал A называется **ноль содержащим интервалом**, если $a < 0 < \overline{a}$.

3.1.3.1.2. Стандартная интервальная арифметика

Арифметические операции над интервальными числами определяются следующим образом. Пусть $* \in \{+, -, \times, /\}$ и A, B ∈ I(**R**), тогда:

$$A * B = \{a * b \mid a \in A, b \in B\},$$

причём в случае деления 0 ∉ B. Нетрудно видеть, что приведённое выше определение соответствует следующим четырём равенствам:

- $A + B = [a, \overline{a}] + [b, \overline{b}] = [a + b, \overline{a} + \overline{b}]$.
- $A - B = [a, \overline{a}] - [b, \overline{b}] = [a - \overline{b}, \overline{a} - b]$.
- $A \times B = [a, \overline{a}] \times [b, \overline{b}] = [\min(ab, \overline{a}b, a\overline{b}, \overline{a}\overline{b}), \max(ab, \overline{a}b, a\overline{b}, \overline{a}\overline{b})]$.
- $A / B = [a, \overline{a}] / [b, \overline{b}] = [a, \overline{a}] / [1/b, 1/\overline{b}]$.

Если A и B — вырожденные интервалы, то приведённые равенства совпадают с обычными арифметическими операциями над действительными числами. Таким образом, интервальное число есть обобщение действительного числа, а интервальная арифметика — обобщение обычной.

Из приведённых равенств непосредственно видно, что интервальные сложение и умножение коммутативны и ассоциативны. Кроме того, в роли нуля и единицы служат вырожденные интервалы [0; 0] и [1; 1].

Более того, можно отметить, что если один из операндов арифметической операции не является вырожденным интервалом, то и результат также есть невырожденный интервал (кроме единственного случая умножения на ноль). Из этого утверждения следует, что для невырожденного интервала A не существует обратных по сложению и умножению интервалов, так как если A + B = 0, AC = 1, то все три интервала должны быть вырожденными в силу сказанного. Таким образом, вычитание не является обратной операцией относительно сложения, а деление не является обратной операцией относительно умножения.

Интересным свойством интервально-арифметических операций является невыполнение закона дистрибутивности (это можно показать на примерах), однако существует интересный закон, называемый законом субдистрибутивности:

$$A(B+C) \subset AB + AC.$$

Кроме того, интервальная арифметика обладает таким важным свойством, как монотонность по включению. Это значит, что если A ⊂ C и B ⊂ D, 0 ∉ D, то:

$$\begin{aligned} &A+B \subset C+D \\ &A-B \subset C-D \\ &AB \subset CD \\ &A/B \subset C/D \end{aligned}.$$

Эти соотношения прямо вытекают из определения интервальных арифметических операций.

3.1.3.2. Обобщения интервальной арифметики

Для некоторых случаев невозможно использование приведённых в предыдущем разделе интервальных арифметических операций. Например, в случае деления одного интервала на другой, второй интервал не должен содержать внутри себя ноль, иначе операция интервального деления не определена. Для того чтобы разрешить такие случаи, был создан ряд обобщений на интервальной арифметике [3].

3.1.3.2.1. Интервальная арифметика с нестандартными вычитанием и делением

В работе [15] вводятся нестандартные интервальные операции вычитания и деления, которые определяются следующим образом:

$$A \div B = \left[\min(a-b, \overline{a}-\overline{b}), \max(a-b, \overline{a}-\overline{b})\right],$$

$$A : B = \begin{cases} \left[\min(a/b, \overline{a}/\overline{b}), \max(a/b, \overline{a}/\overline{b})\right], & \text{если } A, B > 0 \\ \left[\min(a/\overline{b}, \overline{a}/b), \max(a/\overline{b}, \overline{a}/b)\right], & \text{если } A, B < 0 \\ (1/b)A, \ \text{если } 0 \in A, B > 0 \\ (1/\overline{b})A, \ \text{если } 0 \in A, B < 0 \end{cases}.$$

Для интервальной арифметики с такими нестандартными операциями можно выделить ряд интересных свойств (при этом под множеством I*(**R**) подразумевается множество $\{A \mid A \in I(R), 0 \in A\}$):

1. A ÷ B = –(B ÷ A).
2. (–A) ÷ B = A ÷ (–B).
3. Из равенства A ÷ C = B ÷ C не следует, что A = B.
4. Уравнение A + X = B имеет единственное решение: X = B ÷ A.
5. Уравнение A ÷ X = B имеет единственное решение: X = A + B. В случае если ω(A) ≥ ω(B), это уравнение имеет ещё одно решение: X = A + (–B).

6. A ÷ B = A + (–B) тогда и только тогда, когда ω(A) = 0, или ω(B) = 0.

7. A : A = 1, $A \notin I^*(\mathbf{R})$.

8. A : B = $(B : A)^{-1}$, A^{-1} : B = B^{-1} : A.

9. Из равенства A : C = B : C не следует, что A = B.

Для элементов $A \in I^*(\mathbf{R})$ определяется функция ν(A):

$$\nu(A) = \max(\underline{a}/\overline{a}, \overline{a}/\underline{a}).$$

10. Уравнение AX = B имеет решение тогда и только тогда, когда ν(A) ≤ ν(B): X = B : A.

11. Уравнение A : X = B имеет решение: X = AB^{-1}. При ν(A) ≥ ν(B) существует ещё одно решение: X = A : B.

12. Уравнение X : A = B имеет решение: X = AB. При ν(A) ≥ ν(B) существует ещё одно решение: X = A : B^{-1}.

13. A : B = AB^{-1} тогда и только тогда, когда ν(A) = 1, или ν(B) = 1.

Ряд других свойств интервальной арифметики с нестандартными вычитанием и делением можно найти в работе [15].

3.1.3.2.2. Обобщённая интервальная арифметика

Некоторые присущие интервальной арифметике свойства, такие как $A - A \neq 0$, $A / A \neq 1$ и т. д., в ряде случаев приводят к возрастанию ширины получаемых в результате вычислений интервалов. Обобщённая интервальная арифметика позволяет во многих случаях уменьшить влияние этих негативных свойств обычной интервальной арифметики.

Для того чтобы определить обобщённые арифметические операции над интервалами, необходимо определить сами расширенные интервалы. Пусть интервал $X = [\underline{x}, \overline{x}]$ задан в виде x = y + [–c, c], где y = m(X), c = 0.5·v(X) > 0. Таким образом, произвольная точка $x^* \in X$ записывается в виде $x^* = y + x$, $x \in [-c, c]$. Пусть необходимо найти интервал, содержащий множество значений некоторого рационального выражения, зависящего от n переменных, изменяющихся в исходных интервалах $X_1, ..., X_n$. Пусть также переменные $x_i^* \in X_i$ в виде:

$$x_i^* = y_i + x_i,\ x_i \in [-c_i, c_i].$$

Далее любой интервал $X_i^* = Y_i + a_{(r=1,n)} x_r Z_{ir}$, где Y_i, Z_{ir}, i = 1,n — некоторые интервалы, а $x_r \in [-c_r, c_r]$, называется **обобщённым интервалом**. Если положить, что $Y_i = [y_i, y_i]$, $Z_{ii} = [1; 1]$, $Z_{ir} = [0; 0]$, $r \neq i$, $x_r = [-c_r, c_r]$, то $X_i^* = X_i$.

Арифметические операции над обобщёнными интервалами определяются следующим образом:

$$X_i * X_j = Y_k + a_{(r=\overline{1,n})} x_r Z_{kr},$$

где $* \in \{+, -, \times, /\}$. Правила вычисления Y_k и Z_{kr} для каждой интервальной арифметической операции определяются следующим образом:

Сложение:

$$Y_k = Y_i + Y_j,$$

$$Z_{kr} = Z_{ir} + Z_{jr},\ r = \overline{1,n}.$$

Вычитание:

$$Y_k = Y_i - Y_j,$$

$$Z_{kr} = Z_{ir} - Z_{jr},\ r = \overline{1,n}.$$

Умножение:

$$Y_k = Y_i Y_j + a_{(r=\overline{1,n})}\left[0, c_r^2\right] Z_{ir} Z_{jr},$$

$$Z_{kr} = Y_i Z_{jr} + Y_j Z_{ir} + Z_{ir} a_{(s=\overline{1,n}, s\neq r)}\left[-c_s, c_s\right] Z_{is},$$

$$Z_{kr} = Y_i Z_{jr} + Y_j Z_{ir} + \left[-1,1\right]\cdot \left|Z_{ir}\right| \cdot a_{(s=\overline{1,n}, s\neq r)} c_s \cdot \left|Z_{is}\right|.$$

Деление:

$$Y_k = Y_i / Y_j,$$

$$Z_{kr} = \frac{Y_j Z_{ir} - Y_i Z_{jr}}{Y_j \cdot \left(Y_j + \left[-1,1\right] \cdot a_{(s=\overline{1,n})} c_s \cdot \left|Z_{is}\right|\right)}.$$

Нетрудно показать, что если

$$x_i \in X_i = Y_i + a_{(r=\overline{1,n})} x_r Z_{ir},$$

$$x_j \in X_j = Y_j + a_{(r=\overline{1,n})} x_r Z_{jr},$$

то $x_i * x_j \in X_i * X_j$, $* \in \{+, -, \times, /\}$.

Во введённом в работе [2] обобщении интервальных операций допускается деление на интервал, содержащий 0, и наличие ситуации, когда для интервала $[a_1, a_2]$ имеет место выражение $a_2 < a_1$.

Обобщённая интервальная арифметика может применяться для сужения интервалов, содержащих множества значений функции в некоторых случаях. Однако при широких исходных интервалах, на которых задана функция, она зачастую даёт интервалы шире, чем другие способы. При очень узких или вырожденных интервалах лучше использовать обычную интервальную арифметику, так как обобщённая арифметика требует применения большего количества операций, а значит больше машинного времени.

3.1.3.3. Сравнение интервальных величин

Сравнение интервалов определяется только на обычных (не обобщённых) интервалах [40]. Пусть имеется два интервала: $A = [a_1, a_2]$ и $B = [b_1, b_2]$. Если эти два интервала не пересекаются, то можно достоверно утверждать, что один из них больше другого, то есть степень уверенности в выражении «A > B» (или «A < B») равна единице: $\alpha_{A > B} = 1$.

Неопределённость в сравнении возникает тогда, когда два интервала пересекаются. В этом случае невозможно определённо сказать, какой интервал больше, а какой меньше — можно лишь вычислить некоторую степень уверенности в истинности этих утверждений. То есть происходит переход в нечёткозначную логику.

Пусть интервалы A и B пересекаются, и при этом между концами этих интервалов выделены следующие отношения: $b_1 < a_1 < b_2 < a_2$ — то есть интервал A лежит немного правее на области определения. Это ограничение не влияет на общность рассуждений, так как если в решаемой задаче имеется пересечение иного рода, то его легко свести к предлагаемому здесь при помощи декомпозиции и переименования.

Для того чтобы вычислить степень истинности утверждения «A > B» в рассматриваемом случае необходимо воспользоваться следующей формулой [40]:

$$\alpha_{A>B} = \frac{a_2 - b_1}{(b_2 - b_1) + (a_2 - a_1)}.$$

Эта формула основана на предположении о том, что отношение «больше чем» (и другие подобные ему) линейно для интервалов. Это предположение не очень сильное и практически не сужает области применения рассматриваемого метода, кроме как появляется невозможность использования расширенных интервалов.

Вычисленные по приведённой формуле значения степеней истинности в сравнении двух интервалов имеют следующие свойства:

- $\alpha_{A>B} + \alpha_{B>A} = 1$;
- $A \geq B \Rightarrow \alpha_{A>B} \geq \alpha_{B>A}$;
- $(A \geq B) \,\&\, (B \geq C) \Rightarrow (A \geq C) \,\&\, \alpha_{A>C} \geq \max(\alpha_{A>B}, \alpha_{B>C})$;
- $(A \leq B) \,\&\, (B \leq C) \Rightarrow (A \leq C) \,\&\, \alpha_{A<C} \geq \max(\alpha_{A<B}, \alpha_{B<C})$.

Иногда возникает необходимость в сравнении интервала и числа. В этом случае нет необходимости использовать приведённую выше формулу, так как она даёт маловразумительные результаты, даже если число представить в виде вырожденного интервала. В этом случае достаточно воспользоваться следующими формулами:

$$\begin{aligned} \alpha_{A>x} &= \frac{a_2 - x}{a_2 - a_1} \\ \alpha_{x>A} &= \frac{x - a_1}{a_2 - a_1} \end{aligned}.$$

Вполне естественно, что эти формулы работают только в тех случаях, когда число x лежит внутри интервала $A = [a_1, a_2]$. Если число x лежит вне интервала A, то такие объекты сравниваются так же, как и непересекающиеся интервалы.

3.2. Верификация знаний с НЕ-факторами

В этой монографии под процессом верификации знаний понимается логическая, синтаксическая и семантическая проверка продукционных правил. Процесс верификации должен производиться перед процессом вывода, чтобы найти и по возможности исправить (или удалить) следующие типы логических ошибок:

1. *Противоречивые правила* — правила, которые могут приводить к противоречиям при прямом или обратном выводе. При прямом выводе противоречием является получе-

ние двух различных следствий из одной и той же посылки. При обратном выводе противоречие возникает при получении противоречивых предпосылок из одного целевого утверждения.

2. *Неполные правила* — правила, которые теоретически могут участвовать в процессе вывода, но при решении конкретных задач никогда не возникают такие ситуации, когда правило участвует в выводе. Чрезвычайно сложный тип логической ошибки с точки зрения выявления.
3. *Транзитивно-замкнутые (цикличные) правила* — транзитивность правил заключается в проявлении одних и тех же атрибутов в посылке и заключении одного правила. Транзитивная замкнутость (цикличность) проявляется более обширно: если вывод, начатый от одного какого-то атрибута, привёл к тому же самому атрибуту.
4. *Излишние правила* — правила, исключение которых из базы знаний не приводит к неполноте. Такие правила избыточны, так как вывод вполне может обойтись и без таких правил.
5. *Конфликтующие правила* — правила, которые при очередном шаге вывода составляют конфликтующее множество. В большинстве систем машинного вывода этот тип логических ошибок исправляется непосредственно в процессе вывода при помощи тех или иных методов разрешения конфликтов.
6. *Тупиковые правила* — правила, приводящие к таким заключениям, которые одновременно не являются целевыми заключениями, а также не участвуют в дальнейшем выводе. Этот тип логических ошибок может обрабатываться на этапе верификации, но может обрабатываться и в процессе вывода при помощи возвратов при поиске решения.
7. *Несвязанные (изолированные) правила* — правила, которые не участвуют ни в одной цепочке вывода. Атрибуты из посылки таких правил не участвуют в следствиях таких правил и наоборот. Изолированные правила, существующие сами по себе.

Далее рассматривается верификация знаний с выделенными типами НЕ-факторов, однако для каждого НЕ-фактора рассматривается только синтаксическая и семантическая проверка, так как логическая проверка продукций с НЕ-факторами производится абсолютно идентично проверке обычных правил вида «ЕСЛИ – ТО – ИНАЧЕ».

3.2.1. Верификация нечётких знаний

Так как нечёткость присутствует в знаниях обычно в виде нечётких переменных, то есть значений определённых параметров проблемной области, то методы верификации таких знаний в общих чертах схожи с методами верификации достоверных знаний.

Однако в связи с тем, что нечёткость является особым видом НЕ-факторов, существуют дополнительные требования, которые могут накладываться на вид функций принадлежности нечётких переменных при решении тех или иных задач.

В работе [20] сформулирован ряд дополнительных условий, которым в силу своей семантики должны удовлетворять функции принадлежности нечётких множеств, описывающих термы лингвистических переменных.

Пусть $T = \{T_i\}$ ($i \in L = \{1, 2, ..., m\}$) — базовое терм-множество некоторой лингвистической переменной $<\beta, T, X>$, а $<T_i, X, C_i>$ — нечёткая переменная, соответствующая терму $T_i \in T$, $C_i = \{<\mu_{Ci}(x)|x>\}$ ($x \in X$). Пусть X является непрерывным подмножеством оси действительных чисел **R**, это не нарушает общности рассуждения, но позволяет сделать их легче для понимания. Через x_1 и x_2 обозначаются inf X и sup X соответственно, то есть лингвистическая переменная β определена в интервале $[x_1; x_2]$. В этом случае множество T можно упорядочить в соответствии с выражением:

$$(\forall T_i \in T)(\forall T_j \in T)(i > j \leftrightarrow (\exists x \in C_i)(\forall y \in C_j)(x > y)).$$

Это выражение обозначает, что терм, который имеет носитель, расположенный левее на оси действительных чисел **R**, получает меньший номер. Такое упорядоченное терм-множество (можно считать — любой лингвистической переменной) должно удовлетворять следующим условиям:

- $\mu_{C_1}(x_1) = 1, \mu_{C_m}(x_2) = 1$;
- $(\forall T_i \in T \setminus \{T_m\})(0 < \sup\limits_{x \in X} \mu_{C_i \cap C_{i+1}}(x) < 1$;
- $(\forall T_i \in T)(\exists x \in X)(\mu_{C_i}(x) = 1)$;
- $(\forall \beta)(\exists x_1 \in R)(\exists x_2 \in R)((\forall x \in X)(x_1 \le x \le x_2))$.

Таким образом, при верификации нечётких отношений необходимо проверять не только общие свойства продукций, но и частные, связанные с нечёткостью. При верификации значений нечётких переменных может происходить нормализаций функций принадлежности, что приводит к удалению ненормированных данных, то есть своего рода борьба с одним из НЕ-факторов второго типа — ненормированностью.

Остаётся отметить, что при использовании нечёткого логического вывода в процессе верификации нет необходимости в выявлении противоречивых и конфликтующих правил, так как эти логические ошибки разрешаются в процессе вывода.

3.2.2. Верификация знаний с неопределённостью, неточностью и недоопределённостью

С точки зрения верификации знания с неопределённостью, неточностью и недоопределённостью практически ничем не отличаются от обычных знаний. Необходимо отлавливать логические и синтаксические ошибки, связанные с представлением рассматриваемых НЕ-факторов в базах знаний.

Единственная проблема, на которую дополнительно необходимо обращать внимание при верификации знаний с выделенными НЕ-факторами, заключается в том, что интервальные величины могут выходить за границы областей определения этих величин. Естественно, что такие случаи не являются логически правильными, и их также необходимо исправлять на этапе верификации.

Использование неопределённых знаний позволяет избежать необходимости производить верификацию противоречивых знаний, так как в самом механизме вывода на знаниях с неопределённостью заложен метод разрешения конфликтов и противоречий. Однако для неточных недоопределённых знаний верификация противоречивых и конфликтующих правил должна производиться в той же мере, что и для полностью достоверных знаний.

3.3. Преобразование НЕ-факторов из одного вида в другой

Как показывает опыт, при обработке знаний с НЕ-факторами иногда возникает необходимость преобразования одних НЕ-факторов в другие [12]. Наиболее часто такое преобразование необходимо проводить при сопоставлении значений в антецедентах правил во время вывода. Необходимость преобразования возникает в тех случаях, когда сопоставляемые значения имеют различную природу, например нечёткое лингвистическое значение сравнивается с простым числом, либо с числом, к которому приписана некоторая погрешность измерения (неточное значение).

Некоторые НЕ-факторы довольно близки друг к другу по своей природе, например нечёткость и неопределённость в некоторых задачах можно считать одними тем же НЕ-фактором [4] и соответственно обрабатывать при помощи одних и тех же методов. Более того, отдельные виды неточности и недоопределённости совпадают друг с другом с точностью до формализма представления, поэтому для их обработки также разумно пользоваться одним и тем же средством.

Далее рассматриваются наиболее часто встречающиеся на практике случаи преобразования одних НЕ-факторов в другие.

3.3.1. Фаззификация чётких значений

Фаззификация — это процесс приведения «обычных» значений к определённым функциям принадлежности. Необходимость в фаззификации может возникнуть в том случае, если в антецеденте правила, выполнение которого происходит на текущем этапе вывода, встречаются как чёткие, так и нечёткие значения. Такие одновременные «совпадения» могут происходить как в логических, так и в арифметических выражениях, используемых в антецедентах правил.

Алгоритмы фаззификации можно разделить на два вида: преобразование в нечёткие множества данных различных типов (числа, строки и т. д.) и преобразование в нечёткие множества данных с другими НЕ-факторами (неопределённость, неточность). Таким образом, общий процесс фаззификации будет заключаться в комбинировании двух или более методов, представленных далее.

3.3.1.1. Фаззификация чисел

Числа проще всего приводить к нечёткому виду. Для этого достаточно создать дискретную функцию принадлежности, которая имеет значение 1 как раз в фаззифицируемом числе, но на остальной оси чисел такая функция должна иметь значение 0.

Задача фаззификации чисел может возникнуть в случае, если в антецеденте какого-либо правила функция принадлежности некоторого нечёткого множества сравнивается с конкретным числом. В этом случае действительно заданное число достаточно представить в виде дискретной функцией принадлежности и провести операцию сравнения двух функций принадлежности. Однако можно поступить и по-другому, а именно дефаззифицировать функцию принадлежности и сравнить уже два числа. Однако второй способ кажется менее адекватным, так как при помощи него затруднительно вычислить уверенность в полученном результате.

3.3.1.2. Фаззификация строковых атрибутов

Нечёткая переменная — это переменная, принимающая значение в виде функций принадлежности. То есть фактически это пара: «имя переменной — функция принадлежности». При этом имя переменной — это строка. Лингвистическая переменная может быть представлена в виде набора нечётких переменных (то есть терм-множеств). Таким образом видно, что атрибуты строкового типа представляют собой объекты, которые подобны нечётким атрибутам, но за исключением лишь того, что каждому возможному значению нечёткого атрибута приписана определённая функция принадлежности. Последнее утверждение верно только для строковых типов, где количество значений ограничено (перечислимо).

Типы, список значений которых — перечислимое множество заданных строк, можно разделить на два вида: проецируемые на числовую ось и непроецируемые на числовую ось. В первом случае между каждыми двумя строками такого типа можно поставить отношение порядка, то есть узнать, какая строка меньше, а какая больше. В общем случае для такого сравнения можно использовать порядковый номер строки в типе.

В этом случае фаззификация атрибутов, тип которых является строковым и в то же время проецируемым на числовую ось, может проходить абсолютно так же, как и для простых чисел, ведь каждому значению из строкового типа поставлено в соответствие конкретное число. Однако для повышения толерантности нечёткого вывода можно использовать нечёткие числа LR-типа, при этом максимум такого числа будет находиться как раз в числе, соответствующем рассматриваемой строке.

Этот процесс можно рассмотреть на примере. Пусть есть набор строк, представляющий собой строковый тип, проецируемый на числовую ось:

$$T = \{'Аз' \to 1, 'Буки' \to 2, 'Веди' \to 3, 'Глаголь' \to 4, 'Добро' \to 5, 'Есть' \to 6\},$$

а также заданы коэффициенты толерантности, которые представляют собой значения функций принадлежности фаззифицируемой строки на смежных с ней строках, то есть:

$$A = \{0,\ 0.25,\ 1,\ 0.25,\ 0\}.$$

При помощи этих коэффициентов толерантности и числовых значений фаззифицируемых строк можно получать функции принадлежности. Например, для строки «Глаголь» будет создана следующая функция принадлежности:

$$\mu_{Глаголь} = \{1\,|\,0,\ 2\,|\,0,\ 3\,|\,0.25,\ 4\,|\,1,\ 5\,|\,0.25,\ 6\,|\,0\}.$$

Варьируя коэффициенты толерантности, можно добиваться более адекватного представления фаззифицируемых строк для конкретной проблемной области.

Тогда, когда спроецировать строковый тип на числовую ось затруднительно или не представляется возможным, операции сравнения атрибутов и значений в антецедентах правил необходимо проводить на уровне самих строк и имён терм-множеств. Хотя представляется маловероятным существования такой ситуации.

В случае если атрибут принадлежит «чистому» строковому типу, то есть может принимать значение в виде любой возможной строки, фаззификация такого атрибута не представляется возможной, а все операции сравнения должны происходить на уровне символов, например значение атрибута можно сравнивать с именем терм-множества (или, что то же — функции принадлежности).

3.3.1.3. Фаззификация атрибутов, тип которых — перечислимое множество

Атрибуты, имеющие тип перечислимого множества — это более общий случай строковых атрибутов. Поэтому к таким атрибутам вполне можно применять методы, которые относятся к фаззификации строковых атрибутов. Например, к таким атрибутам относятся булевские, принимающие только два значения: «Истина» и «Ложь».

В любом случае, перечислимое множество всегда можно спроецировать на числовую ось (хотя не всегда между элементами такого множества можно провести отношение порядка), поэтому можно использовать алгоритм, использующий коэффициенты толерантности.

3.3.2. Преобразование неопределённости в нечёткость

Иногда необходимо преобразовать неопределённость в нечёткость. Такая задача может возникнуть в случае, если к фаззифицируемому чёткому атрибуту приписаны некоторые коэффициенты уверенности. Как показано в предыдущем разделе фаззификация чётких значений производится при помощи построения функций принадлежности, принимающей значение 1 только в точке на области определения, соответствующей фаззифицируемому чёткому значению.

Если к фаззифицируемому значению приписана степень уверенности в этом значении, то достаточно приравнять значение функции принадлежности на этом значении самой степени уверенности. Проблема возникает лишь в том случае, если степень уверенности представлена в виде интервала, в этом случае возможны два варианта действий:

1. Значение функции принадлежности приравнивается верхней границе интервала уверенности, так как эта граница является и верхним ограничением, выше которого значение уверенности подняться не может.
2. Рассматривается так называемая нечёткость второго порядка, когда значением функции принадлежности в свою очередь является функция принадлежности, областью определения которой и является интервал уверенности.

Второй вариант является более сложным с точки зрения представления и обработки полученных таким образом знаний, а его адекватность рассматриваемым задачам представляется завышенной, поэтому целесообразно использовать первый вариант как довольно точное приближение к реальным решениям.

3.3.3. Методы изменения функций принадлежности в соответствии со степенью уверенности

Неопределённость — НЕ-фактор, который может встречаться вместе с другими НЕ-факторами, в том числе и вместе с нечёткостью. В общем случае это выглядит как приписывание факторов уверенности к значению нечёткого атрибута в антецеденте правила. При вычислении результата логических или арифметических операций таких неопределённо-нечётких значений могут возникнуть некоторые сложности, так как неясно, что потом делать с факторами уверенности, то есть каким образом факторы уверенности влияют на результат логической или арифметической операции.

Для решения этой проблемы можно воспользоваться методом преобразования функции принадлежности в зависимости от приписанного к ней фактора уверенности, который принимает значение из интервала [0, 1]. В этом случае можно считать, что заданный фактор уверенности выражает уверенность в том, что обрабатываемый атрибут описывается именно той функции принадлежности, которая приведена в антецеденте правила.

Необходимо отметить, что если фактор уверенности равен 1, то преобразование функции принадлежности не требуется, так как в этом случае существует полная уверенность в значении нечёткого атрибута. Если фактор уверенности меньше 1, то все точки функции принадлежности необходимо подвергнуть следующему преобразованию:

$$m_j^{'} = m_j k + 0.5(1 - k),$$

где m_j — исходная точка функции принадлежности, m_j’ — получаемая в процессе преобразования точка функции принадлежности, k — фактор уверенности. Приведённая формула производит линейное преобразование значений функции принадлежности и обладает следующими свойствами:

1. m_j’ = m_j при k = 1;
2. m_j’ = 0.5 при k = 0;
3. m_j’ = 0.5 при m_j = 0.5 для любого $k \in [0, 1]$.

На следующем рисунке схематично показан процесс преобразования исходной точки функции принадлежности (m_j) в преобразованную точку (m_j’). При этом видно, что чем ближе фактор уверенности k к значению 0 (то есть «Полная неопределённость»), тем сильнее исходная функция принадлежности превращается в отрезок уровня 0.5.

Вертикальной точечной линией показан коэффициент уверенности, равный 0, наклонной точечной линией — коэффициент уверенности, равный 1, соответственно.

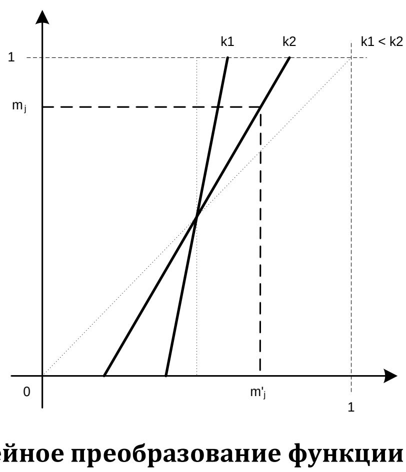

Рисунок 23. Линейное преобразование функции принадлежности

Если к нечёткому значению приписан интервальный коэффициент неопределённости, то возникает такая же проблема, как и при преобразовании неопределённости в нечёткость (см. предыдущий раздел). В этом случае пути решения этой проблемы такие же, как было описано ранее — либо рассматривать только верхнюю границу интервала неопределённости, либо использовать нечёткость второго порядка. Естественно, что с точки зрения представления и обработки нечётких знаний первый способ намного проще, а лёгкость его использования перекрывает адекватность второго метода.

3.3.4. Преобразование неточности в недоопределённость и обратно

Отдельные виды недоопределённости могут быть преобразованы к неточности (и наоборот). Как известно, недоопределённые числовые значения описываются интервалом на оси области определения этих значений. В то же время неточные числовые значения представляют собой точно заданные значения с приписанной к ним абсолютной или относительной погрешностью. То есть для числовых значений неточность представляется в виде:

$$x \pm \Delta x,$$

где Δx — абсолютное значение погрешности измеренной величины. Для получения верхней и нижней границы интервала на числовой оси, в который входит рассматриваемое числовое значение, необходимо воспользоваться формулами:

$$\begin{matrix} a = x - \Delta x \\ b = x + \Delta x \end{matrix},$$

где a и b — нижняя и верхняя границы интервала соответственно. Для обратного преобразования необходимо воспользоваться формулами:

$$\begin{matrix} x = \dfrac{a+b}{2} \\ \Delta x = \dfrac{b-a}{2} \end{matrix}.$$

Необходимо отметить, что такое преобразование неточности в недоопределенность возможно только в случае конечности множеств, которыми задаются недоопределенные значения, иначе это преобразование неверно, так как невозможно задать правило, при помощи которого можно вычислить все элементы бесконечного множества, имея только один выделенный элемент этого множества.

3.3.5. Фаззификация неточности и недоопределённости

Отдельные виды неточности могут быть представлены при помощи интервала. Например, методическая неточность, которая соответствует тем параметрам реального мира, которые получены при помощи каких-либо измерительных приборов. Такие параметры задаются при помощи некоторого конкретного значения и его абсолютной или относительной погрешности, что незатруднительно может быть переведено в интервальный вид. Далее рассматриваются некоторые методы преобразования интервалов на числовых осях в функции принадлежности, определённые на этих же осях.

3.3.5.1. Линейная функция принадлежности уровня α

Можно представить неточные величины таким образом, что внутри интервала функция принадлежности принимает некоторое значение α, а вне интервала — 0. Частным случаем можно считать задание интервальной величины в виде характеристической функции:

$$\begin{aligned} &\lambda(x) = 1,\ x \in [x_1, x_2] \\ &\lambda(x) = 0,\ x \notin [x_1, x_2] \end{aligned}.$$

Однако в данном случае сама неточность величин теряется в процессе нечёткого вывода, так как механизм вывода будет выдавать 1 («Истина») в случае, если интервалы в антецеденте и сукцеденте пересекаются, и 0 («Ложь»), если не пересекаются.

Более адекватным способом представления неточности при помощи функции принадлежности уровня α является такой:

$$\begin{aligned} &\mu(x) = 1 - \delta,\ x \in [x_1, x_2] \\ &\mu(x) = 0,\ x \notin [x_1, x_2] \end{aligned},$$

где δ — относительная погрешность неточной величины.

Здесь при максиминном нечётком выводе более высокая неточность поглотит менее высокую, что зачастую происходит в реальных задачах. Например, если имеется такая задача: «От рулона ткани 20 м ± 40 см отрезается кусок в 3 м ± 5 см. Каков будет результат?» Огрублённо можно принять, что остаток длиной в 17 м сохранит точность в 40 см. Таким образом, если при фаззификации неточности воспользоваться последним предложенным методом, то наблюдается все тот же эффект поглощения менее высокой неточности более высокой.

3.3.5.2. Треугольная функция принадлежности

Другой способ фаззификации неточности, представимой в виде интервала, выглядит следующим образом: внутри интервала выбирается точка (например, середина интервала), в которой значение функции принадлежности полагается равным единице. На границах интервала значение функции принадлежности равно нулю, а от границ к выбранной точке внутри интервала функция изменяется линейно. То есть в случае, если выбрана середина интервала, формулы фаззификации примут такой вид:

$$\begin{aligned} &\mu(x) = 0,\ x \notin [x_1, x_2] \\ &\mu(x) = \frac{(2x - 2x_1)}{x_2 - x_1},\ x \in \left[x_1, \frac{x_1 + x_2}{2}\right] \\ &\mu(x) = \frac{(2x - 2x_2)}{x_1 - x_2},\ x \in \left[\frac{x_1 + x_2}{2}, x_2\right] \end{aligned}.$$

Формулы такого вида описывают треугольные функции принадлежности (то есть один из видов нечётких LR-чисел), причём основанием треугольника является сам интервал, а его вершина находится в выбранной точке. Такая фаззификация более адекватна в случае, если имеется некоторое распределение вероятности нахождения истиной величины параметра реального мира в интервале, которым оценивается этот параметр. Допуская, что значение параметра с наибольшей вероятностью принимает какое-то конкретное значение внутри оценочного интервала (например, середину интервала), треугольные функции принадлежности приведут к результату с менее низкой точностью в случае, если параметры, участвующие в выводе пересекаются малыми областями.

И опять же в данном случае можно варьировать вид треугольника, задавая его высоту так же, как оценку неточности параметра, который оценивается при помощи интервала. То есть можно воспользоваться следующими формулами:

$$\mu(x) = 0,\ x \notin [x_1, x_2]$$
$$\mu(x) = 2(1-\delta)\frac{x - x_1}{x_2 - x_1},\ x \in \left[x_1, \frac{x_1 + x_2}{2}\right].$$
$$\mu(x) = 2(1-\delta)\frac{x - x_2}{x_1 - x_2},\ x \in \left[\frac{x_1 + x_2}{2}, x_2\right]$$

Здесь видно, что чем больше неточность оцениваемого параметра, тем ниже высота треугольника, то есть тем ниже максимум функции принадлежности. Необходимо отметить, что адекватность треугольных функций принадлежности необходимо проверять на практике.

3.3.5.3. Колоколообразная функция принадлежности

Описанные выше треугольные функции принадлежности являются частным случаем колоколообразных функций. Такие функции принадлежности могут задаваться многими способами:

1. **Гауссианы**. Преимущество этого способа состоит в простоте представления (коэффициенты в формуле гауссианы), однако, вычислять точки пересечения гауссиан затруднительно. Гауссианы представляют собой сумму двух сигмоид:

$$sig(\beta, x) = \frac{1}{1 + e^{-\beta x}},$$

где β — коэффициент сигмоиды, от которого зависит степень пологости функции (при помощи этого коэффициента можно варьировать дисперсию колоколообразной функции принадлежности).

2. **Кривые Безье**. Также легко представимые функции (достаточно задать лишь координаты трёх точек и шесть векторов касательных для описания всей колоколообразной функции), но опять же трудно вычислять пересечения.

3. **Кусочно-линейные аппроксимации**. В данном случае необходимо задать список точек, которыми аппроксимируется колоколообразная функция. Чем больше точек задаётся, тем более точно воспроизводится кривая, но в то же время вычислять точки пересечения отрезков прямых чрезвычайно просто.

Точно также варьируя этот метод фаззификации можно задавать высоту колокола в зависимости от точности представляемого параметра, то есть чем ниже точность, тем более приплюснут колокол к числовой оси.

Необходимо отметить, что такой вид представления интервальной неточности подходит в том случае, когда распределение вероятности нахождения параметра изменяется не линейно, а подчиняется нормальному закону распределения, что встречается достаточно часто.

Соответственно при фаззификации недоопределённых значений можно привести эти недоопределённые значения к виду неточных и фаззифицировать их каким-либо из методов, предложенных для фаззификации неточности.

Заключение

Как было показано, к настоящему времени наиболее полно проработана математическая теория, предоставляющая механизмы для всесторонней обработки нечёткости. Эта ситуация вполне правомерна, так как сам этот НЕ-фактор был введён в рассмотрение ещё в 50-ых годах XX века. С того времени было создано несколько сотен тысяч научных публикаций, в основном приходящихся на японских и европейских учёных. Эти публикации раскрывают все возможные нюансы нечёткой математики во всех её аспектах.

Менее разработанной оказывается теория обработки неопределённости, хотя и относительно этого НЕ-фактора написано довольно большое количество публикаций. Это связано с тем, что многие исследователи считают неопределённость одним из видов нечёткости, что позволяет в какой-то мере применять и те аппараты, которые разрабатывались для работы с нечёткостью. Кроме того, неопределённость можно обрабатывать различными вероятностными и статистическими методами (что показывают соответствующие методы обработки, полученные из расширения классических вероятностных методов), поэтому здесь также можно учитывать огромное количество разработок в области теории вероятности и математической статистики.

Самыми малоисследованными НЕ-факторами из выделенных оказываются неточность и недоопределённость. И если математические теории для работы с неточностью более или менее разрабатывались, так как интервальная арифметика необходима во многих прикладных областях человеческой деятельности, то недоопределённость как отдельный НЕ-фактор была выделена совсем недавно, что дополнительно сказывается на количестве публикаций на тему работы с недоопределённостью.

Следующая таблица обобщает полученные в процессе исследования результаты относительно существующих подходов, методов и технологий извлечения, представления и обработки знаний с НЕ-факторами.

Таблица 10. Количество публикаций, раскрывающих аспекты работы с НЕ-факторами

	Приобретение		Представление	Обработка			
	Эксперт	БД		Арифметика	Сравнение	Вывод	Верификация
Нечёткость	+ + +	+	+ + +	+ +	+	+ + +	+
Неопределённость	+	?	+	—	—	+ +	?
Неточность	+	?	+	+	+	+	?
Недоопределённость	+	?	+	+	+	+	?

Представленная таблица качественно раскрывает общее количество публикаций и работ в той или иной области, рассматривающей работу с НЕ-факторами (нечёткостью, неопределённостью, неточностью и недоопределённостью). Каждая ячейка таблицы обозначает публикации на определённую тему. Например, ячейка на пересечении столбца «Представление» со строкой «Неточность» показывает количество публикаций на тему представления неточных знаний в системах, основанных на знаниях.

Знак «+» обозначает, что публикации на заданную тему существуют, а количество знаков «+» отображает сравнительное количество таких публикаций: «+ + +» — довольно обширное поле публикаций, «+ +» — вполне разработанная тема, «+» — существуют небольшое количество статей и работ по рассматриваемой теме.

Знак «—» обозначает отсутствие публикаций на эту тему вследствие некорректно сформулированной темы исследований. Знак «?» обозначает тот факт, что либо работы на рассматриваемую тему отсутствуют вовсе, либо их количество слишком мало, поэтому в процессе подготовки этой монографии такие работы найдены не были.

Таким образом видно, что наибольшее количество белых пятен в науке о НЕ-факторах располагается в области изучения неточности и недоопределённости. Это также связано с тем, что эти НЕ-факторы были выделены совсем недавно (в рамках последнего десятилетия), поэтому все исследования до сих пор остаются в большей мере на уровне теоретических изысканий. Кроме того, некоторые научные школы даже отвергают существование любых НЕ-факторов, за исключением нечёткости, а всю обработку недостоверных знаний возлагают на проработанные за многие годы математические формализмы и аппараты работы с нечёткостью.

Список литературы

На русском языке

1. **Аверкин А. Н.** *Построение нечётких моделей мира для планирования в условиях неопределённости* // В кн.: Семиотические модели при управлении большими системами. — М.: АН СССР, 1979.
2. **Аленфельд Г., Херцбергер Ю.** *Введение в интервальные вычисления*. — М: Мир, 1987.
3. **Алтунин А. Е., Семухин М. В.** *Модели и алгоритмы принятия решений в нечётких условиях*: Монография. — Тюмень: Издательство Тюменского государственного университета, 2000.
4. **Батыршин И. З.** *Методы представления и обработки нечёткой информации в интеллектуальных системах* // Новости искусственного интеллекта. — М.: АИИ, 1996. № 2.
5. **Берг А. И., Бирюков Б. В., Геллер Е. С., Поваров Г. Н.** *Управление, информация, интеллект*. — М.: Мысль, 1976.
6. **Блишун А. Ф.** *Сравнительный анализ методов измерения нечёткости* // Изв. АН СССР. Техническая кибернентика. 1988. № 5. — с. 152–175.
7. **Блохнин А. Н.** *Нечёткий вывод, использующий преобразование функций принадлежности* // В кн.: Известия РАН. Теория и системы управления. 1997, № 5. — с. 119-124.
8. **Борисов А. Н., Крумберг О. А., Фёдоров И. П.** *Принятие решений на основе нечётких моделей*: Примеры использования. — Рига : Зинатне, 1990.
9. **Броневич А. Г., Каркищенко А. Г.** *Теория нечётких мер и обобщения байесовской схемы классификации статистических данных* // В кн: Перспективные информационные технологии и интеллектуальные системы / Электронный журнал: http://pitis.tsure.ru/
10. **Городецкий В. И.** *Моделирование недоопределённых знаний* // В кн.: SCM'98. Международная конференция по мягким вычислениям и измерениям. Сборник докладов. В 2-х томах. — Санкт-Петербург: АОЗТ «කописервис», Том 1. — с. 98-102.
11. **Городецкий В. И., Тулупьев А. Л.** *Формирование непротиворечивых баз знаний с неопределённостью* // Изв. Академии Наук. Теория и системы управления. 1997. № 5.
12. **Душкин Р. В., Рыбина Г. В.** *Об одном подходе к автоматизированному извлечению, представлению и обработке знаний с НЕ-факторами* // В кн.: Известия РАН. Теория и системы управления. 1999, № 5. — с. 34-44.
13. **Жуковин В. Е., Оганесян Н. А., Бурштейн Ф. В., Корелов Э. С.** *Об одном подходе к задачам принятия решений с позиции теории нечётких множеств* // В кн.: Методы принятия решений в условиях неопределённости. — Рига: РПИ, 1980. — с. 12-16.
14. Искусственный интеллект в 3-х кн. Кн. 1. *Системы общения и экспертные системы*. Справочник / Под ред. Э. В. Попова. — М.: Радио и связь, 1990.

15. **Калмыков С. А., Шокин Ю. И., Юлдашев З. Х.** *Методы интервального анализа*. — Новосибирск: Наука, 1986.
16. **Каменский В. С.** *Методы и модели неметрического шкалирования (обзор)*. — Автоматика и телемеханика, 1977, № 8. — с. 118-152.
17. **Кнорринг В. Г.** *Предложения по терминологии в области шкал* // В кн.: SCM'98. Международная конференция по мягким вычислениям и измерениям. Сборник докладов. В 2-х томах. — Санкт-Петербург: АОЗТ «Кописервис», Том 2. — с. 63-69.
18. **Круглов В. В., Дли М. И., Голунов Р. Ю.** *Нечёткая логика и искусственные нейронные сети*: Учебное пособие. — М.: Издательство Физико-математической литературы, 2001.
19. **Кудинов Ю. И.** *Нечёткие модели вывода в экспертных системах* // В кн.: Известия РАН. Теория и системы управления. 1997, № 5. — с. 75-83.
20. **Мелихов А. И., Бершгейн Л. С., Коровин С. Я.** *Ситуационные советующие системы с нечёткой логикой*. — М.: Наука. Гл. ред. физ.-мат. лит., 1990.
21. **Нариньяни А. С.** *Недоопределённость в системах представления и обработки знаний* // Изв. АН СССР. Техн. кибернетика. 1986. № 5.
22. **Нариньяни А. С.** *НЕ-факторы и инженерия знаний: от наивной формализации к естественной прагматике* // В кн.: Сборник трудов IV национальной конференции по Искусственному Интеллекту. Т. 1 (КИИ-94, Рыбинск, сентябрь 1994 г.). — с. 9-18.
23. *Нечёткие множества в моделях управления и искусственного интеллекта* // Под ред. Поспелова Д. А. М.: Наука, 1986.
24. **Покатаева Е. Н.** *Чёрное золото 90-х годов, или как видеть мир глазами другого человека* // Вычислительная техника и её применение, № 10. — М.: Знание, 1990.
25. **Попов Э. В., Фоминых И. Б., Кисель Е. Б., Шапот М. Д.** *Статические и динамические экспертные системы*: Учеб. Пособие. — М.: Финансы и статистика, 1996.
26. *Прикладные нечёткие системы* / Под ред. Тэрано Т., Асаи К., Сугено М. — М.: Мир, 1993.
27. **Робинсон Дж.** *Машинно-ориентированная логика, основанная на принципе резолюций* // В кн.: Кибернетический сборник, вып. 7, 1970. — с. 194-218.
28. **Рыбина Г. В.** *Проектирование систем, основанных на знаниях*: Учебное пособие. — М.: МИФИ, 2000.
29. **Рыбина Г. В., Душкин Р. В.** *Некоторые аспекты автоматизированного извлечения и обработки знаний с НЕ-факторами* // В кн.: SCM'99. Международная конференция по мягким вычислениям и измерениям. Сборник докладов. В 2-х томах. Санкт-Петербург: АОЗТ «Кописервис», Том 2. — с. 105-108.
30. **Рыбина Г. В., Душкин Р. В.** *НЕ-факторы: лингвистические аспекты извлечения* // В кн.: Труды Международного семинара Диалог-2002 по компьютерной лингвистике и её приложениям в двух томах. — Под ред. А. С. Нариньяни. Т. 2. — М.: НАУКА, 2002.

31. **Тарасов В. Б.** *Анализ и моделирование НЕ-факторов на полярных шкалах* // В кн.: Интегрированные модели и мягкие вычисления в искусственном интеллекте: Сборник трудов Международного научно-практического семинара, Коломна, 17-18 мая 2001. — М.: Наука. Физматлит, 2001. — с. 65-71.

32. **Франселла Ф.**, **Баннистер Д.** *Новый метод исследования личности*. — М.: Прогресс, 1987.

33. **Чень Ч.**, **Ли Р.** *Математическая логика и автоматическое доказательство теорем*. — М.: Мир, 1983.

34. **Шапот М. Д.** *Вывод решений в условиях неопределённости в системе ЭКО* // Экспертные системы на персональных компьютерах. Материалы семинара. — М.: МДНТМ, 1989.

35. **Шер А. П.** *Согласование нечётких экспертных оценок и функция принадлежности в методе размытых множеств* // В кн.: Моделирование и исследование систем автоматического управления. — Владивосток: ДВНЦ АН СССР, 1978. — с. 111-118.

36. **Элти Дж.**, **Кумбс М.** *Экспертные системы: концепции и примеры* / Пер. с англ. Б. И. Шитикова. — М.: Финансы и статистика, 1987.

На английском языке

37. **Averkin A. N.** *Fuzzy Logic Simulation Technology in General Strategy of Intelligent System Designing* // Proceedings of the Second International Conference on Application of Fuzzy Systems and Soft Computing, edited by R. A. Aliev, K. W. Bonfig, F. Aliev, F. Wieland, ICAFS'96, Siegen, Germany — June 25-27, 1996.

38. **Dadgostar A. S.** *A decentralized reactive fuzzy scheduling system for cellular manufacturing systems*. — PhD Thesis, University of South Wales, Australia, 1996.

39. **Dempster A. P.** *Upper and Lower Probabilities Induced by a Multi-valued Mapping* // Annals of Mathematical Statistics 38, 1967.

40. **Dorohonceanu B.**, **Marin B.** *A simple method for comparing fuzzy numbers*. — CAIP Center, Rutgers University, Piscataway, NJ 08854-8088.

41. **Dubois D.**, **Prade H.** *Fuzzy sets and systems*. — Academic Press, vol. 144, New York, 1980.

42. **Jaszkiewicz A.** *Multiple objective metaheuristic algorithms for combinatorial optimization*. — Habilitation thesis, 360. — Poznan University of Technology, Poznan, 2001.

43. *Knowledge Discovery Through Data Mining: What Is Knowledge Discovery?* — Tandem Computers Inc., 1996.

44. **Li M.**, **Vitanyi P.** *An introduction to Kolmogorov complexity and its applications*. — Springer-Verlag, N. Y., 1997.

45. **Makoto O.**, **Hitoshi M.**, **Kenichiro S.**, **Masaaki O.** *Self-Tuning of Fuzzy Reasoning by Brent's Method* // Proceedings of 15th IMACS World Congress of Scientific Computation, Modelling and Applied Mathematics. — August 24-29, 1997, Berlin, Germany. — Wissenschaft & Technik Verlag, V. 4.

46. **Mendel J.** *Uncertain rule-based fuzzy logic systems: introduction and new directions*. — Prentice-Hall, Upper Saddle River, NJ, 2001.

47. **Moor R. E.** *A servey of interval methods for differential equations* // Proc. Of 23rd IEEE Conference. Decision and Control. — Las Vegas, Nevada, 1984, v. 3, New York, 1984, p. 1529-1535.

48. **Nguyen H. T.**, **Kreinovich V.** *On efficient representation of expert knowledge by fuzzy logic*. — CS, University of Texas at El Paso, 2002.

49. **Osgood C. E.**, **Suci G. J.**, **Tannenbaum P. H.** *The measurement of meaning*. — Univercity of Illinois Press, Urbana, 1957. — pp. 1-342.

50. **Saaty T. L.** *Exploring the interface between hierarchies, multiple objectives and fuzzy sets*. — Fuzzy Sets and Systems, 1978, V. 1. — pp. 57-69.

51. **Schwandt H.** *An interval arithmetic approach for the constraction of an almost globally convergent method for the solution of the nonlinear poisson equation on the unit square*. — SIAM J. Sci. a St. Comput. — 1984, v. 5, № 2. — p. 427-452.

52. **Shafer G.** *A Mathematical Theory of Evidence*. — Princeton University Press, Princeton, NJ, 1976.

53. **Skala H. J.** *On any-valued logics, fuzzy sets, fuzzy logics and their applications*. — Fuzzy Sets and Systems, 1978, V. 1. — pp. 129-149.

54. **Tanaka K.**, **Mizumoto M.** *Fuzzy programs and their execution* // In: Fuzzy Sets and Their Applications to Cognitive and Decision Processes / Ed. By L. A. Zadeh et al. — New York: Academic Press, 1975. — pp. 41-76.

55. **Wilson N.**, **Moral S.** *Fast Markov Chain Algorithms for Calculating Dempster-Shafers Belief* // Proceedings of 12th European Conference on Artificial Intelligence. — August 11-16, 1996, Budapest, Hungary. — John Wiley & Sons Ltd.

56. **Zadeh L. A.** *Fuzzy logic, neural networks and soft computing*. — Communications of the ACM, 37:3, 1994.

57. **Zadeh L. A.** *Fuzzy Sets* // Information and Control. — V. 8, pp. 338-353.

Принимаются благодарности

Вниманию всех читателей! Данная книга издана в электронном виде и распространяется абсолютно бесплатно. Вы можете свободно использовать её для чтения, копировать её для друзей, размещать в библиотеках на сайтах в сети Интернет, рассылать по электронной почте и при помощи иных средств передачи информации. Вы можете использовать текст книги частично или полностью в своих работах при условии размещения ссылок на оригинал и должном цитировании.

При этом автор будет несказанно рад получить читательскую благодарность, которая позволит как улучшить текст данной книги, так и более качественно подойти к подготовке следующих книг. Благодарности принимаются на счёт в платёжной системе «Яндекс.Деньги», на который также можно перечислить малую лепту и при помощи терминалов:

4100137733052

Убедительная просьба; по возможности, при перечислении благодарности указывать в пояснении к переводу наименование книги или какое-либо иное указание на то, за что именно выражается благодарность.

ДУШКИН Роман Викторович

Методы получения, представления и обработки знаний с НЕ-факторами

Главный редактор —
Корректор —
Вёрстка *Душкин Р. В.*
Дизайн обложки —

Распространяется бесплатно

Москва
2011