

О манипулировании знаниями с использованием нечетких множеств

В.Б.Новосельцев, Е.А.Романчук
(Томский государственный университет)

Аннотация

В работе обсуждаются некоторые возможности применения аппарата нечетких логик к анализу и манипулированию формализованными знаниями (с учетом свойств и связей между существенными компонентами). Предлагается некоторая модификация (по отношению к классической) машины вывода (inference engine), реализующей взаимодействие конкурирующих объектов и учитывающей статику и динамику поведения «знаний». Предлагаемый формализм одновременно обеспечивает корректную работу как с четкими, так и с нечеткими (fuzzy) характеристиками элементов базы знаний, что позволяет использовать предлагаемый подход при реализации "семантически-ориентированных" комплексов манипулирования знаниями.

1. Нечеткое множество

Определим *домен* как *множество* (однородных) элементов. Критерий принадлежности некоторого элемента e конкретному домену D задается одноместным предикатом $D(e)$, определенным на универсуме элементов U . Важно отметить, что имя предиката и имя домена – суть одно и то же. Двухместным предикатом $is_a(e, D)$, определенным на декартовом произведении $U \times 2^U$ (универсумов элементов и доменов), задается отношение «*быть частью*» – очевидно, что $is_a(e, D) \equiv D(e)$. Теперь, сопоставив *ИСТИНЕ* единицу (1), а *ЛЖИ* ноль (0), традиционным способом можно определить характеристическую функцию (функцию принадлежности элемента домену) – $m(e, D)$ с областью определения $U \times 2^U$, так что $D \equiv \{e \in U | m(e, D) = 1\}$. Традиционно вводится и понятие *поддомена* D' домена D : $D' \equiv \{e \in D | m(e, D') = 1\}$. – Если принадлежность элемента множеству несущественна или очевидна из контекста, соответствующий фрагмент в формальной записи может опускаться, с учетом сказанного и для обозначения характеристической функции (для фиксированного домена D) может использоваться запись $m_D(e)$ или $m(e)$.

Распространив область значений функции $m()$ на весь отрезок $[0, 1]$, т.е. отказавшись от ее дискретности, мы переходим к нечетким понятиям. При этом значение характеристической функции определяет уже *степень принадлежности* элемента множеству и *нечеткое вхождение* одного множества в другое. В дальнейшем вместо символа «|» с традиционной теоретико-множественной трактовкой часто будет использоваться наклонная черта «/», а «.../m(...)» будет означать «...со степенью (принадлежности) m(...)».

В такой трактовке функция $m()$, являясь (качественной) оценкой принадлежности элемента множеству, с одной стороны задает нечеткое множество, а с другой – определяет понятие нечеткого подмножества. Данные выше определения (с учетом соглашений)

принимают вид: $D \equiv \{e \in U | m_D(e)\}$ (или $D \equiv \{e/m(e)\}$) и, для $D' \subseteq D$, $D' \equiv \{e \in D | m_{D'}(e)\}$ (или $D' \equiv \{e/m_{D'}(e)\}$).

Может возникнуть вопрос: принадлежит ли элемент нечеткому множеству, если степень его принадлежности равна нулю? – Это особенно актуально при сравнении нечетких множеств на сходство-подобие. Остановимся на двух вариантах:

(1) Формирование множества путем явного поэлементного перечисления – $D \equiv \{e_1/m(e_1), \dots, e_k/m(e_k), \dots\}$ (очевидно, это не *нечеткое* множество, элементы принадлежат ему *по определению* независимо от значения характеристической функции).

(2) Введение понятия *носителя (carrier)* C_D – это множество элементов универсума, для которых характеристическая функция принадлежности определяемому множеству D строго больше нуля – $C_D \equiv \{e \in U | m_D(e) > 0\}$. Понятие подмножества $D' \subseteq D$ при этом переопределяется следующим образом: $D' \equiv \{e \in C_D | m_{D'}(e)\}$.

2. Оценка подобия

Отношением подобия в теории нечетких множеств называется любое бинарное отношение, заданное на некотором множестве E и обладающее свойствами рефлексивности и симметричности; если к тому же оно обладает свойством транзитивности, то это уже *отношение сходства*. Аналогом этого отношения в классической теории множеств служит отношение эквивалентности, но это достаточно «сильное» отношение.

Построим оценку подобия *нечетких подмножеств* A и B «четкого множества» E следующим образом: $I(A, B) = 1 - \rho(A, B)$, где $\rho(A, B)$ есть относительное расстояние Хэмминга либо относительное евклидово расстояние.

Доказательство симметричности этой оценки вытекает из симметричности (по определению) расстояния $\rho(A, B)$. Рефлексивность же легко выводится из антирефлексивности и ограниченности $\rho(A, B)$. Таким образом, мы действительно имеем оценку подобия *нечетких подмножеств* A и B «четкого множества» E .

3. Нечеткие и лингвистические переменные. Основные понятия

Достаточно нетривиальной является задача трансляции значения *лингвистической переменной* в нечеткое подмножество базового множества для данной переменной, а уж тем более сложна задача обратной трансляции. – Здесь возможно наложить несколько ограничений на вид функции принадлежности. К примеру, функция принадлежности должна иметь вид «колокольчика» в середине базового множества, и вид «полу-колокольчика» на концах. Подобное ограничение допустимо, если не будут использоваться фразы типа: «не маленький» или «маленький или очень большой». Тогда задача трансляции «нечеткое множество – фраза» и обратная ей существенно упрощаются.

Примем определения и обозначения, касающиеся нечетких и лингвистических переменных, введенные в [2, 5], незначительно их пополнив.

Определение 1: *Объектом* предметной области называется любой факт, предмет или событие X предметной области удовлетворяющее «четкому» предикату $is(X)$ (для текущего состояния предметной области).

Определение 2: *Контекстом объекта* или *множеством нечетких атрибутов* назовем совокупность нечетких переменных, если для каждой из них помимо наименования задано еще и имя лингвистической переменной значением которой они могут являться.

Пример 2. Зададим контекст объекта - предприятия.

Определение 3: Пусть P – прямое произведение n множеств и M – его множество принадлежностей; *нечеткое n -арное отношение* E определяется как нечеткое подмножество P , принимающее свои значения в M .

Определение 4: Связью между объектами A и B назовем элемент нечеткого бинарного отношения E , определенного на множестве объектов-знаков и принимающего свои значения в $M=[0,1]$, обладающий следующими свойствами:

а) Каждая связь имеет функцию доступа от первого элемента ко второму, называемую *ролью* – имя бинарного отношения, другими словами семантика отношения.

б) Каждая связь имеет вес – скаляр из отрезка $[0,1]$ – множества принадлежности отношения E , описывающий возможность существования данного отношения между объектами A и B .

Определение 5: *Текущим состоянием схемы* назовем упорядоченную совокупность объектов, всех контекстов для данного объекта и бинарных отношений между парами объектов, на которых определены связи в фиксированный момент времени. Текущее состояние схемы может быть выражено с помощью нечеткого упорядоченного множества, элементами которого являются \langle Имя объекта/1 \rangle , элементы контекстов как нечеткие множества, \langle Роль_связи (Объект1, Объект2)/Вес_связи \rangle .

В силу упорядоченности семантика подобного множества сохранится, даже если в каждом его элементе отбросить первую часть, получив при этом однородный вектор со значениями элементов принадлежащими $M=[0,1]$.

Теперь мы имеем право говорить о *расстоянии между состояниями схемы* как о расстоянии между нечеткими упорядоченными множествами или векторами.

4. Динамика

В свое время Заде выдвинул идею о том, что нечеткая переменная (в контексте нечеткого подмножества) может, в определенных случаях, рассматриваться как распределение возможностей или субъективных вероятностей. Мы, в свою очередь, любое событие предметной области (совершенное или прогнозируемое) представляем в виде нечеткой переменной, то есть в *качестве* распределения субъективных вероятностей. Таким образом, представляется возможным построение событийного аппарата на механизмах сопоставления и выборки нечетких множеств-событий. Для ознакомления с существующими методами нечеткого логического вывода на сопоставлении можно обратиться к соответствующему обзору в [6]. В данной работе немного пересмотрена концепция многоуровневого выбора (см. [2,7]) и структура вида:

if EVENT... then ... else_if ... then ... else_if ... fi ... fi ... fi

естественным образом заменяется на

case EVENT of value₁:{...};... value_k:{...};...end_case.

Как легко заметить, мы избавились от вторичного сравнения (для *else*), поскольку в нечеткой логике тавтология состоятельна (см., например, великолепную работу Кофмана [1] о свойствах подобия и сходства на нечетких множествах). Одной из основных проблем остается проблема приведения двух нечетких подмножеств к одному четкому надмножеству. Известно несколько методов подобного приведения ([3,4]) – остановимся на следующих:

1. Элиминирование несуществующих элементов (как крайний вариант).
2. Спряmlение или аппроксимация (в зависимости от эвристики) на основе учета характеристик *крайних* или *всех* элементов множества. – К сожалению, подобный метод подходит лишь для упорядоченных множеств.

Теперь возможно описать принципиальную схему работы событийного аппарата. Она (схема) основана на следующих положениях:

1. Любое событие, совершенное или прогнозируемое, представляется нечетким множеством (распределение субъективных вероятностей).
2. Любое событие имеет тип (хотя возможно использование иерархии типов событий с верхним абстрактным бестиповым событием) и так называемый нечеткий контекст – совокупность нечетких типизированных множеств – отвечающий за классификацию события. Событие при этом описывается тройкой $\langle \text{СОБЫТИЕ}, \text{ТИП}, \text{КОНТЕКСТ} \rangle$.
3. Любой *объект* системы (контейнер), имеют множество четких или нечетких контекстов, а также набор обрабатываемых событий с *методами* реагирования на них.
4. Любая реакция *объекта* порождает новое событие, которое, ставится в очередь событий, и, возможно, новое действие, связанное с модификацией системы в целом или ее частей.
5. Для упорядоченной пары $\langle \text{СОБЫТИЕ}, \text{КОНТЕКСТ} \rangle$ определена бинарная операция: $(\mathbf{A}, \mathbf{B}) \otimes (\mathbf{C}, \mathbf{D}) \rightarrow \mathbf{M} \times [(\mathbf{A} \times [\mathbf{B} \# \mathbf{D}]) \# \mathbf{C}]$, где (\mathbf{A}, \mathbf{B}) – реальное или прогнозируемое событие, \mathbf{C} – предопределенный идеал события для объекта, \mathbf{D} – контекст объекта, операция « \times » – умножение на скаляр, а « $\#$ » – алгебраическая мера подобия. При этом в качестве \mathbf{M} выступает событие, порожденное реакцией объекта на событие (\mathbf{A}, \mathbf{B}) .

Вследствие все того же различия между сравниваемыми множествами, представляется разумным иметь различные алгоритмы сравнения нечетких множеств с четкими же, нечетких с четкими и четких с четкими. Поскольку в любом случае нужна алгебраическая (замкнутая) оценка подобия, предлагается использовать нечто аналогичное ядерной оценке для "размывания" четких значений при сравнении. Для случая непрерывного ранжированного множества применяется, во-первых, его дискретизация на основе эвристик, определяемых предметной областью, а затем – сравнение по тем же алгоритмам.

Подчеркнем еще раз, что основной целью предлагаемого подхода является поддержка «динамизма» схемы базы знаний. – Любой объект способен (или нет) реагировать на сообщения (выражающие внутренние или внешние для системы события), предлагаемые ему системой. В случае внешнего события используется циркулярная рассылка сообщений. Внутренние сообщения рассылаются как циркулярно, так и селективно. Подобным же образом с помощью сообщений в системе реализуются любые действия по модификации экстенционала.

Некоторым отходом от объектно-ориентированного стандарта событийного механизма является реализация этого механизма в виде совокупности событий, посылаемых системе «одновременно». Основная проблема в работе с сообщениями заключается в том, что моделируется недетерминированный процесс «изменения» экстенционала базы знаний на детерминированном вычислителе. При этом, во-первых, к одному объекту могут быть адресованы несколько сообщений, и все они должны найти объект в таком виде, в каком он находился на момент их генерации (вне зависимости от очередности поступления на обработку); во-вторых, объект, созданный в результате обработки сообщения, не должен отзываться на сообщения уровня создателя. – Преодоление этих препятствий достигается путем введения очередей сообщений, регулируемых «темпоральными» семафорами. Очередь сообщений поделена на уровни, и каждому уровню соответствует «такт времени», в который было сгенерировано сообщение. Чем больше уровень сообщения, тем позже оно было получено. Соответственно, переход к следующему уровню осуществляется только по исчерпанию сообщений текущего уровня.

На основе предложенной модели реализован макет программного комплекса, иллюстрирующий ее основные положения и предполагающий дальнейшее развитие.

Прекрасной инструментальной средой при этом выступают функциональные комплексы на базе языка Лисп (например, Allegro CL [7]).

Литература

1. Кофман А. Введение в теорию нечетких множеств. М.: Радио и связь, 1977 – 432 с.
2. Мелихов А.С., Бернштейн Л.С., Коровин С.Я. Ситуационные советующие системы с нечеткой логикой. М.: Наука, 1990 – 271 с.
3. Поспелов Д.А. Логико-лингвистические модели в системах управления. М.: Энергоиздат, 1981 – 231 с.
4. Дюбуа Д., Прад А. Теория возможностей. Приложения к представлению знаний в информатике. М.: Радио и связь, 1990 – 288 с.
5. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. М.: Мир, 1976 – 162 с.
6. Ягер Р. Нечеткие множества и теория возможностей, последние достижения. М.: Радио и связь, 1986 – 406 с.
7. www.franz.com