

К.т.н. Т.В. Зайцева, Е.В. Нестерова, к.с.н. С.В. Игрунова,
Н.П. Путивцева, О.П. Пусная (НИУ «БелГУ»)

T. Zaitseva, E. Nesterova, S. Igrunova, N. Putivzeva,
O. Pusnaya

АЛГОРИТМ ПЕРЕВОДА КОЭФФИЦИЕНТОВ НЕЧЕТКОЙ
ЛОГИКИ В КОЭФФИЦИЕНТЫ УВЕРЕННОСТИ
ПРИ РАЗРАБОТКЕ ЭКСПЕРТНОЙ СИСТЕМЫ В СРЕДЕ GURU

ALGORITHM OF TRANSFER OF COEFFICIENTS OF FUZZY LOGIC
IN CONFIDENCE COEFFICIENTS BY DEVELOPMENT
OF EXPERT SYSTEM IN THE ENVIRONMENT OF GURU

В статье рассмотрены два взаимодополняющих компонента термина нечеткость – ненадежность и неопределенность. Рассмотрена возможность использования теории вероятности при выводе в условиях неопределенности. Приведен пример алгоритма преобразования нечеткого множества в единственное значение.

In article two complementary components of the term unsharpness – unreliability and uncertainty are considered. Possibility of usage of the probability theory is considered at an output in the conditions of uncertainty. The example of algorithm of conversion of unsharp set in a unique value is resulted.

Ключевые слова: нечеткость, неопределенность, ненадежность, коэффициент уверенности, алгоритм.

Key words: unsharpness, uncertainty, unreliability, coefficient of confidence, algorithm.

Экспертные системы (ЭС) применяются для решения слабо-структурированных и неструктурированных задач. Одной из характерных особенностей таких задач является необходимость учета факторов, которые не могут быть известны достоверно. Типичные примеры таких факторов - различные прогнозы, субъективные оценки решений и т.д. Кроме того, задачи, решаемые с использованием ЭС, обычно имеют много вариантов решения, каждый из которых может быть рекомендован с определенной степенью уверенности. В связи с этим одним из обязательных требований к ЭС является наличие средств представления и обработки нечетких и недостоверных знаний [1-3].

Смысл термина *нечеткость* неоднозначен. Существует

достаточно большое количество взаимодополняющих компонентов данного понятия. недетерминированность выводов; многозначность; ненадежность, неполнота; неточность и др. Так, например, ненадежность знаний означает, что для оценки их достоверности нельзя применить двухбалльную шкалу (1 — абсолютно достоверные; 0 — недостоверные знания). Поэтому широкое применение на практике получили нечеткие выводы, строящиеся на базе нечеткой логики, которая ведет свое происхождение от теории нечетких множеств. Однако для более тонкой оценки достоверности знаний лучше использовать вероятностный подход, основанный на теореме Байеса. Рассмотрим возможность перехода от коэффициентов нечеткой логики к вероятностным коэффициентам уверенности.

Основное отличие нечеткой логики от классической заключается в том, что она интерпретируется не только значениями «истина» и «ложь», но и промежуточными значениями.

Обобщенную формулу нечеткого множества можно представить в виде [3]:

$$B = \sum_{i=1}^n \frac{x_i}{\mu(x_i)}, \quad (1)$$

где: B — нечеткое множество;

$\mu(x)$ — функция принадлежности;

x_i — i -е значение базовой шкалы

Функция принадлежности определяет субъективную степень уверенности эксперта в том, что данное конкретное значение базовой шкалы соответствует определяемому нечеткому множеству.

Для представления операций с нечеткими знаниями существует много способов, как правило, все они являются эвристиками. В качестве примера можно привести стандартные определения в нечеткой логике (логике Заде) [4]:

Истинно (не x) = $1.0 - \text{истинно}(x)$;

Истинно (x и y) = минимум (истинно (x), истинно (y)),

Истинно (x или y) = максимум (истинно (x), истинно (y))

Данные записи отличаются от определений в вероятностном подходе [3,5], например,

ИЛИ (x, y) = $x + y - x * y$

Нечеткое правило логического вывода представляет собой упорядоченную пару (A, B), где A — нечеткое подмножество про-

странства входных значений X , B — нечеткое подмножество пространства выходных значений Y . Как правило, оно имеют вид:

если цена велика и спрос низкий, то оборот мал,

где цена и спрос — входные переменные; оборот — выходное значение; велика, низкий и мал — функции принадлежности (нечеткие множества), определенные на множествах значений цены, спроса и оборота, соответственно.

Нечеткие правила вывода образуют базу правил. Следует отметить, что в нечеткой управляющей системе, отличие от традиционной, работают все правила одновременно, однако степень их влияния на выход различна [5]. Принцип вычисления суперпозиции многих влияний на окончательный результат лежит в основе нечетких управляющих систем. Процесс обработки нечетких правил вывода в управляющей системе в среде GURU состоит из четырех этапов:

1) Вычисление степени истинности левых частей правил (между «если» и «то») — определение степени принадлежности входных значений нечетким подмножествам, указанным в левой части правил вывода.

2) Модификация нечетких подмножеств, указанных в правой части правил вывода (после «то»), в соответствии со значениями истинности, полученными на первом этапе.

3) Объединение (суперпозиция) модифицированных подмножеств.

4) Скаляризация результата суперпозиции — переход от нечетких подмножеств к скалярным значениям.

Для определения степени истинности левой части каждого правила нечеткая управляющая система вычисляет значения функций принадлежности нечетких подмножеств от соответствующих значений входных переменных. Например, определение степени вхождения конкретного значения переменной цена в нечеткое подмножество велика, т.е. истинность предиката «цена велика». Полученное значение истинности используется для модификации нечеткого множества, указанного в правой части правила. Для выполнения такой модификации используют один из двух методов: «минимум» (correlation-min encoding) или «произведение» (correlation-product encoding).

Первый метод ограничивает функцию принадлежности для

множества, указанного в правой части правила, значением истинности левой части. Во втором методе значение истинности левой части используется как коэффициент, на который умножаются значения функции принадлежности.

Результат выполнения правила — ассоциирование переменной и функции принадлежности, указанной в правой части правила, т.е. нечеткое множество.

Выходы всех правил вычисляются нечеткой управляющей системой отдельно, однако в правой части нескольких из них может быть указана одна и та же нечеткая переменная. При определении обобщенного результата необходимо учитывать все правила, для этого система производит суперпозицию нечетких множеств. Эта операция называется нечетким объединением правил вывода. Например, правая часть правил

если цена мала, то спрос велик

если цена велика, то спрос мал

содержит одну и ту же переменную — спрос. Два нечетких подмножества, получаемые при выполнении этих правил, должны быть объединены управляющей системой.

Как правило, суперпозиция функций принадлежности нечетких множеств определяется как их объединение. Другой метод суперпозиции состоит в суммировании значений всех функций принадлежности. Самым простым, но и наименее часто используемым, является подход, когда суперпозиция не производится. Выбирается одно из правил вывода, результат которого используется в качестве интегрального результата.

Конечный этап обработки базы правил вывода — переход от нечетких значений к конкретным скалярным. Процесс преобразования нечеткого множества в единственное значение называется «скаляризацией».

При переходе от нечеткого вывода к четкому определенному значению могут использоваться различные способы: метод центра тяжести; метод первого максимума; метод среднего максимума; метод наименьшего максимума [3]. Однако, при учете неопределенности знаний и наличии априорной информации необходимо искать другие способы.

Рассмотрим одну из компонент нечеткости — неопределенность. Неопределенность может быть двух типов:

1. неопределенность в истинности самой посылки (например, если степень уверенности в том, что A истинно составляет 90%, то какие значения примет B);

2. неопределенность самого правила (например, можно сказать, что в большинстве случаев, но не всегда, если есть A , то есть и B).

Еще более сложная ситуация возникает в случае, если правило имеет вид: *если (A и B), то (C)*, тогда можно быть в некоторой степени уверенным как в истинности каждой из посылок (A , B), а тем более их совместного проявления, так и в истинности самого вывода.

Таким образом, существуют четыре проблемы, которые возникают при проектировании и создании систем с неопределенными знаниями:

1. Как количественно выразить степень определенности при установлении истинности (или ложности) некоторой части данных?

2. Как выразить степень поддержки заключения конкретной посылкой?

3. Как использовать совместно две (или более) посылки, независимо влияющие на заключение?

4. Как быть в ситуации, когда нужно обсудить цепочку вывода для подтверждения заключения в условиях неопределенности?

Рассмотрим возможности использования теории вероятности при выводе в условиях неопределенности, т.е. байесовское оценивание. Правило Байеса имеет вид

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)}. \quad (2)$$

Это равенство является основой для использования теории вероятности в управлении неопределенностью, т.к. оно обеспечивает путь для получения условной вероятности события B при условии события A . Данное соотношение позволяет управлять неопределенностью и «делать вывод как вперед, так и назад» [4]

Неопределенность можно промоделировать, приписывая утверждениям некоторые характеристики, отличные от «истина» или «ложь». Степень уверенности может выражаться в форме действительного числа, заключенного в некотором интервале, например между -1 и $+1$. Таковую числовую характеристику называют по-разному: коэффициент определенности, коэффициент уверенности

степень доверия. Границы интервала обозначают следующее:

- +1 — система в чем-то полностью уверена;
- 0 — у системы нет знаний об обсуждаемой величине;
- 1 — высказанная посылка или заключение полностью неверно.

Промежуточные величины отражают степень доверия или недоверия к указанным ситуациям. Коэффициент неопределенности в выводимом заключении представляет собой число, которое должно объединить и отразить источники внутренних ошибок.

Приведем схему приближенных рассуждений для реализации в среде GURU. Основной вычислительный прием, который используется для нахождения коэффициента определенности, сводится к следующему:

$$t(\text{заключение}) = ct(\text{посылка}) * ct(\text{импликация}).$$

Для импликации вида имеет место

$$\text{ЕСЛИ } (e1 \text{ и } e2) \text{ ТО } (c)$$

Согласно оценке, сделанной в GURU, коэффициент определенности посылки равен коэффициенту определенности наименее надежной из посылок, т.е.

$$ct(e1 \text{ и } e2) = \min(ct(e1), ct(e2)).$$

Для

$$\text{ЕСЛИ } (e1 \text{ или } e2) \text{ ТО } (c)$$

общее правило комбинирования, по которому вычисляется коэффициент определенности посылки, заключается в том, что коэффициент определенности дизъюнкции равен коэффициенту ее сильнейшей части, т.е.

$$ct(e1 \text{ или } e2) = \max(ct(e1), ct(e2)).$$

В случае, когда используются два правила для поддержки, например:

$$\text{ЕСЛИ } (e1) \text{ ТО } (c) \quad ct(\text{заключения})=0.9$$

$$\text{ЕСЛИ } (e2) \text{ ТО } (c) \quad ct(\text{заключения})=0.8$$

Был предложен следующий механизм:

$$Stotal = \text{коэффициент определенности из правила 1}$$

$$+ - \text{коэффициент определенности из правила 2}$$

$$- - (\text{коэффициент определенности из правила 1})$$

$$* - (\text{коэффициент определенности из правила 2})$$

При вычислении итоговых коэффициентов необходимо учитывать знаки:

- 1) если оба коэффициента определенности положительны:
 $ctotal = ct1 + ct2 - ct1 * ct2;$
- 2) если оба коэффициента определенности отрицательны:
 $ctotal = ct1 + ct2 + ct1 * ct2;$
- 3) когда отрицателен один коэффициент:
 $ctotal = (ct1 + ct2) / (1 - \min(\text{abs}(ct1), \text{abs}(ct2)));$
- 4) когда одна определенность равна +1, а другая -1:
 $ctotal = 0.$

Вычисление коэффициентов уверенности заключения имеет явно выраженный модульный характер, поскольку не нужно принимать во внимание никакой иной информации, кроме той, что имеется в данном правиле. При этом не имеет никакого значения, как именно получены коэффициенты уверенности, характеризующие исходные данные.

При построении ЭС часто используется эта особенность. Полагается, что для всех правил, имеющих дело с определенным параметром, предпосылки каждого правила логически независимы [4]. Пусть, например, существует зависимость между признаками $e1$ и $e2$. Не следует распределять по двум правилам

ЕСЛИ ($e1$) ТО (c) $ct(\text{заключения})$

ЕСЛИ ($e2$) ТО (c) $ct(\text{заключения})$

Следует сгруппировать их в единое правило

ЕСЛИ ($e1$ и $e2$), ТО (c) $ct(\text{заключения})$.

В основе этой рекомендации лежит одно из следствий теории вероятностей, гласящее, что $P(H | E1, E2)$ не может быть простой функцией от $P(H | E1)$ и $P(H | E2)$.

Выражения для условной вероятности не могут в этом смысле рассматриваться как модульные.

Выражение $P(B | A) = t$ не позволяет заключить, что $P(B) = t$ при наличии A , если только A не является единственным известным признаком.

Если кроме A мы располагаем еще и знанием E , то нужно сначала вычислить $P(B | A, E)$, а уже потом можно будет что-нибудь сказать и о значении $P(B)$.

Такая чувствительность к контексту может стать основой очень мощного механизма логического вывода, но за это придется платить существенным повышением сложности вычислений.

Литература

1. Нейлор К. Как построить свою экспертную систему. М., Энергоатомиздат, 1991. 286 с.
2. Романов, В.П. Интеллектуальные информационные системы в экономике. Под ред. Н.П. Тихомирова. М., Экзамен, 2003. 496 с.
3. Рыбина Г.В., Душкин Р.В., Демидов Д.В. Модели и методы обработки недостоверных знаний в инструментальном комплексе АТ-ТЕХНОЛОГИЯ. – В сб. Труды II-го Междунар. НПС. Коломна, 2003, с. 401-407.
4. Жиравок А. Н. Нечеткие множества и их использование при принятии решений. – "Соросовский образовательный журнал", 2001, т. 7.
5. Корневский Н.А., Крупчатников Р.А., Горбатенко С.А. Синтез нечетких сетевых моделей, обучаемых по структуре данных для медицинских экспертных систем. – "Медицинская техника", 2008, № 2, с. 18-24.

Статья поступила 19 05 2011

К.т.н. Т.В. Зайцева, Р.М. Устинов, О.П. Пусная
(НИУ «БелГУ»)

T.V. Zaitseva, R.M. Ustinov, O.P. Pusnaya

**КОМПЬЮТЕРНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА ОБРАБОТКИ
СТАТИСТИЧЕСКИХ ДАННЫХ С УЧЕТОМ ВЕРОЯТНОСТНОЙ
НЕОПРЕДЕЛЕННОСТИ КЛАССИФИКАЦИИ**

**COMPUTER REALIZATION OF ALGORITHM OF PROCESSING
OF THE STATISTICAL DATA TAKING INTO ACCOUNT
LIKELIHOOD UNCERTAINTY OF CLASSIFICATION**

В статье рассмотрена стратегия принятия решения с учетом вероятностной неопределенности классификации параметров. Представлена программная реализация алгоритма обработки статистических данных с учетом апостериорных условных вероятностей.

The article discusses the strategy decision, taking into account the probabilistic uncertainty of classification parameters. Presented a software implementation of the algorithm of statistical data, taking into account a posteriori conditional probabilities.

Ключевые слова: статистические данные, априорные условные вероятности, алгоритм