УДК 519.816:004.032.2 **doi: 10.20998/2522-9052.2020.4.07**

В. Г. Левашенко [1], А. С. Ляшенко [2], Г. А. Кучук [3]

[1] Жилинский университет, Жилина, Словакия
[2] Харьковский национальный университет радиоэлектроники, Харьков, Украина
[3] Национальный технический университет “ХПИ”, Харьков, Украина

ПОСТРОЕНИЕ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ НА ОСНОВЕ НЕЧЕТКИХ ДАННЫХ

Аннотация. Разработка инструментария оценки принимаемых решений является актуальной и востребованной задачей на современном этапе развития информационных технологий. Таким инструментарием являются, например, системы поддержки принятия решений (СППР). В работе предлагается математический аппарат построения СППР. Построение СППР предполагает анализ имеющихся результатов наблюдений или измерений и выработку стратегии проверок исходных параметров в виде дерева нечетких решений или продукционных правил. Основу предлагаемого аппарата составляют суммарные информационные оценки (информация и энтропия) для нечетких наборов данных. Использование нечетких данных наиболее полно соответствует человеческой природе, поскольку на практике люди часто применяют субъективные ощущения и априорные знания, чем точные вероятностные критерии. Поэтому, используя нечеткую логику и рассматривая степень возможности как нечеткую меру, эксперты имеют возможность описывать реальные данные с достаточной точностью. Исследована взаимосвязь предложенных суммарных информационных оценок. В работе приведены примеры, демонстрирующие использование предлагаемого математического аппарата на практической задаче. В дальнейшей работе, авторы планируют привести результаты экспериментальных исследований предлагаемого подхода и его сопоставление с иными известными методами и алгоритмами. Указанное сопоставление представляется для широкого круга формализованных данных, хранящихся в известном репозитории *UCI Machine Learning Repository*. В качестве сопоставляемых методов и алгоритмов планируется выбрать иные алгоритмы построения деревьев нечетких решений, алгоритмы байесовской классификации, построения деревьев решений C4.5, CART и метод ближайших соседей.

Ключевые слова: системы поддержки принятия решений; нечеткая логика; деревья нечетких решений.

Введение

Постановка проблемы. Происходящие в мире процессы, характеризуются динамичностью, неопределенностью и сложностью прогнозирования их развития. Принимаемые на различных уровнях решения зачастую противоречивы и не позволяют достичь заранее поставленных целей. Поэтому актуальной задачей является разработка инструментария оценки принимаемых решений. Таким инструментарием являются, например, системы поддержки принятия решений (СППР). Практическое применение этих систем позволяет в минимальные сроки и с минимальными затратами выполнить анализ текущего состояния объекта или процесса, выработать оптимальное по ряду критериев решение об его развитии, исследовать чувствительность принятого решения и дать научно обоснованные выводы о результатах принятого решения. Таким образом, окажется возможным минимизировать затраты на развитие объекта и потери от недостаточно проработанных решений, что является одним из основных приоритетов развития на современном этапе.

В настоящее время вопросы создания методов и алгоритмов, реализуемых в СППР, достаточно хорошо изучены. Однако, используемые в современных СППР методы и алгоритмы должны учитывать возможную не стохастическую неопределенность исходных данных, вызванную недостаточной точностью измерений, неполным соответствием методик оценки исходных показателей их реальному состоянию, неоднозначностью восприятия человеком (экспертом) окружающей действительности. Одним из перспективных подходов к учету такой неопределенности является представление исходных значений нечеткими и многозначными данными. Применение в современных СППР данных такого типа позволяет более адекватно описывать объекты и анализировать их состояние. Поэтому актуальной задачей развития современных СППР является разработка инструментария нечеткого вывода для анализа нечетких и многозначных данных.

Цель работы заключается в описании математического аппарата построения СППР, а именно модуля логического вывода. Предлагаемая СППР предполагает генерацию последовательности проверок исходных параметров, используя информацию о предыдущих результатах проверок. Таким образом, оказывается возможным осуществить прогнозирование возможных ситуаций с заданной степенью достоверности и с минимальными затратами. Применение предлагаемого математического аппарата показано на простейших примерах.

Постановка задачи принятия решений

Как правило, любой объект или процесс описывается системой сложных и взаимосвязанных показателей. Между этими показателями существует функциональная зависимость. Одним из способов представления такой зависимости является структурная функция. В этом случае исходные показатели объекта интерпретируются как исходные атрибуты структурной функции, а результирующие – как ее результирующие атрибуты. Значения, которые принимают показатели объекта, есть значения соответствующих атрибутов структурной функции [1].

Поскольку реальный объект имеет сложную структуру, то система описывающих его показателей достаточно громоздка и в результате структурная функция обладает значительной вычислительной сложностью. Одним из подходов к снижению этой сложности является декомпозиция структурной функции. Суть декомпозиции заключается в разбиении этой функции на подфункции с меньшим числом атрибутов.

Аргументы таких подфункций будем интерпретировать как промежуточные атрибуты, а сами подфункции также являются структурными функциями меньшей сложности и, в свою очередь, могут быть декомпозируемы. Будем считать, что структурные функции являются функциями *m*-значной логики. На рис.1 представлена иерархия структурных функций и функций *m*-значной логики [2].

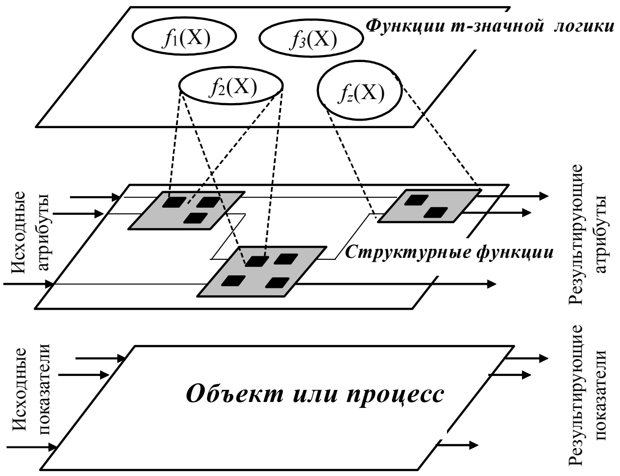

Рис. 1. Представление объекта в виде структурной и логической функций

Основной составляющей процесса принятия решений является выбор решения из множества альтернатив. В ряде случаев анализ исходной ситуации и выбор наилучшего решения осуществляется путем сопоставления с решениями, принятыми в прошлом [3-5]. При этом следует минимизировать затраты на анализ исходной ситуации, за счет определения последовательности наиболее значимых показателей, описывающих ситуацию. Такой выбор формулируется следующим образом. Имеется исходное пространство (среда) решения, описываемое независимыми показателями и конечное множество вариантов решения (альтернатив). Требуется построить модель выбора альтернативы, лучшей в некотором конкретном смысле в зависимости от конкретного состояния независимых показателей, описывающих среду решения.

Таким образом, принятие решения предполагает выбор наилучшего решения из совокупности возможных альтернатив. Этот выбор осуществляется на основе анализа значений исходных показателей. Под термином исходный показатель процедуры принятия решения понимаем измеряемую характеристику (свойство) объекта или процесса, оказывающую влияющие на принимаемое решение.

Пусть имеется n исходных показателей. Каждый i-й (i=1,...,n) исходный показатель ассоциирован с исходным атрибутом A_i, описываемым i-й лингвистической переменной, принимающей m_i ($m_i \geq 2$) возможных значений (термов) $A_{i,1},\ldots,A_{i,j},\ldots,A_{i,m_i}$. В свою очередь решение ассоциировано с результирующим атрибутом, описываемым лингвистической переменной B, а имеющиеся m_b ($m_b \geq 2$) альтернатив решения – значениями этой переменной $B_1,\ldots,B_{jb},\ldots,B_{m_b}$.

Тогда измерение (оценка) i-го исходного показателя предполагает установление всех m_i значений $A_{i,j}$ атрибута A_i. А принятие решения предполагает установление всех m_b значений результирующего атрибута B. Наилучшим решением будет та jb-я альтернатива (jb=1,...,m_b), которой соответствует максимальное значение B_{jb} результирующего атрибута B.

Пример 1. Для принятия решения о надежности возврата кредита оцениваются 4 исходных показателя: *Возраст клиента*, *Годовой доход*, *Стаж работы* и *Покрытие кредита залогом*, ассоциированные с соответствующими исходными атрибутами A_1, A_2, A_3 и A_4. Результатом является оценка надежности возврата кредита. Все исходные и результирующий атрибуты описываются лингвистическими переменными (табл. 1).

Таблица 1 – **Сопоставление показателей и ассоциированных с ними атрибутов, описываемых лингвистическими переменными**

Показатель	Атрибут	Значения атрибута
Тип атрибута – исходный		
Возраст клиента	*Возраст* (A_1)	*Молодой* ($A_{1,1}$); *Средний* ($A_{1,2}$); *Пожилой* ($A_{1,3}$)
Годовой доход	*Доход* (A_2)	*Высокий* ($A_{2,1}$); *Средний*($A_{2,2}$); *Низкий* ($A_{2,3}$)
Стаж работы	*Стаж* (A_3)	*Малый* ($A_{3,1}$); *Большой* ($A_{3,2}$)
Покрытие залогом	*Залог* (A_4)	*Недостаточный* ($A_{4,1}$); *Достаточный* ($A_{4,2}$)
Тип атрибута – результирующий		
Надежность возврата кредита	*Надежность* (B)	*Полная* (B_1); *Частичная* (B_2); *Низкая* (B_3)

Таким образом, исходные A_i (i = 1,...,n) и результирующий B атрибуты описываются лингвистическими переменными и определены группами термов $\{A_{i,1},\ldots,A_{i,j},\ldots,A_{i,m_i}\}$ и $\{B_1,\ldots,B_{jb},\ldots,B_{m_b}\}$ соответственно.

Нечеткая логика и деревья решений

Эксперт, анализируя события реального мира, быстро и с высокой точностью и надежностью оценивает полученную информацию. В своей деятельности эксперт часто оперирует такими оценками, как “*уверен*”, “*возможно*”, “*абсолютно невозможно*” и др. В результате оказывается допустимым представить эмпирические характеристики посредством числовых значений. Использование для представления алфавита четкой (булевой или многозначной) логик не всегда адекватно описывает эти характеристики, а в ряде случаев не позволяет достичь необходимой точности. Сложность описания состоит в некорректной трансформации числовых значений в лингвистические на границах разных термов. Так, небольшое изменение исходного действительного значения на границе приводит к получению иного значения лингвистического терма.

Общая проблема использования четких данных – их неполное соответствие человеческой природе. На практике люди применяют собственные субъективные ощущения, априорные знания и личную память больше, чем точные вероятностные критерии распределения различных данных. Поэтому, используя нечеткую логику и рассматривая степень возможности как нечеткую меру, эксперты описывают реальные данные с достаточной точностью. При этом принадлежность элементов к множеству задается функцией принадлежности.

Определение 1. Нечеткий набор A, принадлежащий множеству U характеризуется *функцией принадлежности* μ_A: $U \rightarrow [0,1]$. Для каждого элемента u в U $\mu_A(u)$ задает меру принадлежности элемента u к A.

Для $u \in$U равенство $\mu_A(u)=1$ означает, что u однозначно принадлежит аргументу A и $\mu_A(u) = 0$ означает, что u не принадлежит аргументу A, и $0 < \mu_A(u) < 1$ означает, что u частично принадлежит A. В случае, если $\mu_A(u) = 0$ или $\mu_A(u) = 1$ для всех $u \in U$, мы имеем дело с четкими данными.

Пример 2. Поясним на рис. 2 взаимосвязь действительных и лингвистических значений, представленных в многозначной (m = 3) и нечеткой формах. Действительному числу x* соответствует значение *Молодой* лингвистической переменной с уверенностью 0.7 и значение *Средний* с уверенностью 0.3.

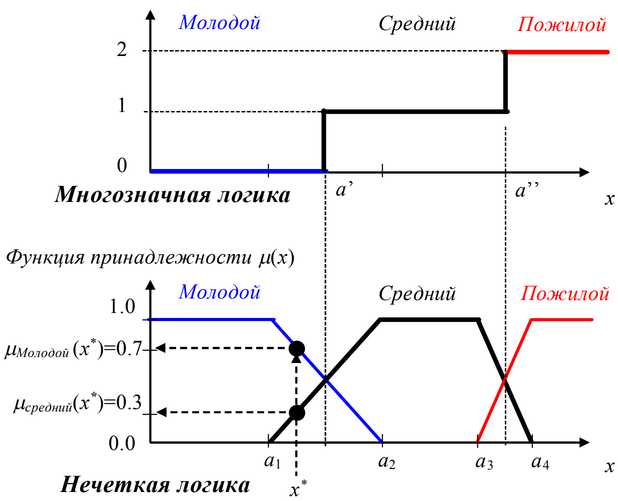

Рис. 2. Пояснение взаимосвязи действительных и лингвистических значений, представленных в многозначной (m = 3) и нечеткой формах

В целом, если значение x немного меньше величины a', то мы вынуждены представлять ее термом *Молодой*, а если немного больше, то уже совершенно иным термом *Средний.* При использовании нечеткой логики этой граничной проблемы удается избежать. Поскольку при подобном увеличении значения x функции принадлежности термов *Молодой* и *Средний* изменятся незначительно.

Вопрос получения лингвистических значений не является предметом исследований данной работы. Авторы рекомендуют ознакомиться с анализом и алгоритмом, представленные в работах [6, 7].

Определение 2. Мощность (мера *cardinality*) нечеткого набора A равна $\mathrm{M}(A)=\Sigma_{u\in U}\, \mu_A(u)$.

Будем полагать, что исходными данными для принятия решений является таблица с нечеткими данными, содержащая формализованные сведения о N возможных ситуациях (комбинаций значений исходных атрибутов) и рекомендуемой в этих ситуациях альтернативе решения [8, 9]. Эти рекомендации априори получены, например, на основе наблюдений, экспертных оценок или путем преобразования действительных значений в лингвистические переменные.

Определение 3. *Таблица с нечеткими данными* есть шестерка {CS, CV, AS, AV, P}, где CS– множество исходных атрибутов, анализируемых при принятии решения; CV – множество значений исходных атрибутов; AS – множество результирующих атрибутов (в дальнейшем, для упрощения изложения будем полагать наличие единственного результирующего атрибута); AV – множество значений результирующего атрибута; P – множество из N ситуаций, представляемое наборами значений исходных и результирующего атрибутов.

Пример 3. Априори заданные результаты N=16 измерений значений 4 исходных (*Возраст*, *Доход*, *Стаж*, *Залог*) и результирующего (*Надежность*) атрибутов представлены в табл. 2. Наименование лингвистических переменных и их значений поясняется в примере 1 (табл. 1).

Оценка значения атрибута представляется действительным числом в диапазоне [0; 1] таким образом, чтобы сумма оценок каждого атрибута была равна 1. Так, например, оценка показателя *Возраст* = = {*Молодой*, *Средний, Пожилой*} = {0.9, 0.1, 0.0} интерпретируется как оценка *Молодой* с уверенностью 90% и *Средний* с уверенностью 10%.

Одной из задач, решаемых при функционировании интеллектуальной системы, является формирование правил классификация исходных данных. Анализ публикаций показал перспективность использования для задач классификации деревьев решений. Рассмотрим дерево, в котором каждая ветвь связывает начальный и конечный узлы. Узел, ветви которого не имеют начальных узлов, называется корнем, а узлы, все ветви которого не имеют конечных узлов – листья.

Определение 4. Пусть $A_i \subset \mathbf{A}$ $(1 \le i \le n)$ n групп нечетких атрибутов, где $|A_i|$ >1. Дерево нечетких решений (*Fuzzy decision tree*) – направленное дерево, удовлетворяющее условиям [8]:

- каждый узел дерева описывается атрибутом A_i из $\mathbf{A}$;
- каждый лист соответствует одному или нескольким значениям классификационного решения;
- каждая группа нечетких наборов A_i, соответствует атрибуту, и каждый нечеткий набор есть значение этого атрибута $A_{i,j}$.

Таблица 2 – **Априори заданные результаты 16 наблюдений, представленные нечеткими данными**

№	***Возраст*** (A_1)			***Доход*** (A_2)			***Стаж*** (A_3)		***Залог*** (A_4)		***Надежность*** (B)		
	$A_{1,1}$	$A_{1,2}$	$A_{1,3}$	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$	$A_{3,1}$	$A_{3,2}$	$A_{4,1}$	$A_{4,2}$	B_1	B_2	B_3
1.	0.9	0.1	0.0	1.0	0.0	0.0	0.8	0.2	0.4	0.6	0.0	0.8	0.2
2.	0.8	0.2	0.0	0.6	0.4	0.0	0.0	1.0	0.0	1.0	0.6	0.4	0.0
3.	0.0	0.7	0.3	0.8	0.2	0.0	0.1	0.9	0.2	0.8	0.3	0.6	0.1
4.	0.2	0.7	0.1	0.3	0.7	0.0	0.2	0.8	0.3	0.7	0.9	0.1	0.0
5.	0.0	0.1	0.9	0.7	0.3	0.0	0.5	0.5	0.5	0.5	0.0	0.0	1.0
6.	0.0	0.7	0.3	0.0	0.3	0.7	0.7	0.3	0.4	0.6	0.2	0.0	0.8
7.	0.0	0.3	0.7	0.0	0.0	1.0	0.0	1.0	0.1	0.9	0.0	0.0	1.0
8.	0.0	1.0	0.0	0.0	0.2	0.8	0.2	0.8	0.0	1.0	0.7	0.0	0.3
9.	1.0	0.0	0.0	1.0	0.0	0.0	0.6	0.4	0.7	0.3	0.2	0.8	0.0
10.	0.9	0.1	0.0	0.0	0.3	0.7	0.0	1.0	0.9	0.1	0.0	0.3	0.7
11.	0.7	0.3	0.0	1.0	0.0	0.0	1.0	0.0	0.2	0.8	0.3	0.7	0.0
12.	0.2	0.6	0.2	0.0	1.0	0.0	0.3	0.7	0.3	0.7	0.7	0.2	0.1
13.	0.9	0.1	0.0	0.2	0.8	0.0	0.1	0.9	1.0	0.0	0.0	0.0	1.0
14.	0.0	0.9	0.1	0.0	0.9	0.1	0.1	0.9	0.7	0.3	0.0	0.0	1.0
15.	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.8	0.2	0.0	0.0	1.0
16.	1.0	0.0	0.0	0.5	0.5	0.0	0.0	1.0	0.0	1.0	0.5	0.5	0.0
У	**6.6**	**5.8**	**3.6**	**6.1**	**5.6**	**4.3**	**5.6**	**10.4**	**6.5**	**9.5**	**4.4**	**4.4**	**7.2**

Один из первых алгоритмов для построения деревьев четких решений ID3 путем обучения на основе начальной выборки, предложен в работе J.R. Quinlan [3]. В основе алгоритма лежит использование информационных оценок. Дальнейшим развитием этого алгоритма являются алгоритмы C4.5, CART и др.

Существует несколько подходов к обобщению ID3 алгоритма на область нечетких данных [8-13]. Основная разница между этими подходами заключается в использовании различных эвристик при выборе атрибутов, которые следует ассоциировать с узлом, и окончание процесса ветвления в листе. Эти подходы основаны на энтропии Luca de Termini [8, 9], совокупная взаимная информация (CMI) [10], приближенных вычислениях (Rough set) [12] и др.

Анализ использования алгоритмов FDT при решении практических задач представлен в [13-15]. Эти работы показывают, что их применение в классификационные задачи позволяют получить хорошие результаты. В этом виде задач, цель FDT – предсказать значение цели переменная из значений нескольких исходных атрибутов. Результат классификация обычно определяется по множеству листьев. FDT может быть легко преобразован в набор правил нечеткой классификации. Пример такого преобразования приведен в работе [16].

Информационно-теоретический подход

Все комбинации значений нечетких атрибутов возникают с определенными частотами, описываемых функцией принадлежности. Анализ этих частот позволяет получить важные свойства атрибутов, такие как собственная и совместная информация, условная и взаимная информация, а также их энтропию [17].

Информация в нечетких данных. Предлагаемые информационные оценки для нечетких данных, позволяют оценить влияние значений атрибутов друг на друга.

Определение 5. *Собственная* информация в значении $A_{i,j}$ нечеткого атрибута A_i вычисляется как

$$\mathrm{I}(A_{i,j}) = \log N - \log\ M(A_{i,j}) = \log N + \mathbf{I}(A_{i,j})\,, \quad (1)$$

где здесь и далее log – логарифм по основанию 2; ($i = 1,\ldots,n$; $j = 1,\ldots,m_i$), $\mathbf{I}(A_{i,j})$ *суммарная собственная* информация в значении $A_{i,j}$ нечеткого атрибута A_i :

$$\mathbf{I}(A_{i,j}) = -\log\ M(A_{i,j}).$$

Определение 6. *Совместная* информация в значениях $A_{i2,j2}$ и $A_{i1,j1}$ нечетких атрибутов A_{i2} и A_{i1} вычисляется по правилу

$$\begin{aligned}\mathrm{I}\left(A_{i1,,j1}, A_{i2,j2}\right) &= \log N - \log\ M\left(A_{i1,j1} \times A_{i2,j2}\right) = \\ &= \log N + \mathbf{I}\left(A_{i1,j1}, A_{i2,j2}\right),\end{aligned} \quad (2)$$

где $\mathbf{I}(A_{i1,j1},A_{i2,j2})$ *суммарная совместная* информация в значениях $A_{i2,j2}$ и $A_{i1,j1}$ нечетких атрибутов A_{i2} и A_{i1}:

$$\mathbf{I}\left(A_{i1,j1}, A_{i2,j2}\right) = -\log \mathrm{M}\left(A_{i1,j1} \times A_{i2,j2}\right).$$

Определение 7. *Условная* информация описывает неопределенность в значении $A_{i2,j2}$ нечеткого атрибута A_{i2} в случае, если значение $A_{i1,j1}$ другого нечеткого атрибута A_{i1} уже стало известным:

$$\begin{aligned}\left(A_{i2,j2} \mid A_{i1,j1}\right) &= \mathrm{I}\left(A_{i1,j1}, A_{i2,j2}\right) - \mathrm{I}\left(A_{i1,j1}\right) = \\ &= -\ \log \mathrm{M}\left(A_{i2,j2} \times A_{i1,j1}\right) + \log \mathrm{M}\left(A_{i1,j1}\right) = \\ &= \mathbf{I}\left(A_{i2,j2} \mid A_{i1,j1}\right),\end{aligned} \quad (3)$$

где $\mathbf{I}(A_{i1,j1}|A_{i2,j2})$ – *суммарная условная* информация в значениях $A_{i2,j2}$ и $A_{i1,j1}$.

Определение 8. *Взаимная* информация между значениями $A_{i2,j2}$ и $A_{i1,j1}$ нечетких атрибутов A_{i2} и A_{i1} используется в качестве оценки степени зависимости значения $A_{i2,j2}$ относительно значения $A_{i1,j1}$ и обратно

$$\begin{aligned}&\mathrm{I}\left(A_{i1,j1}; A_{i2,j2}\right) = \\ &= \mathrm{I}\left(A_{i1,j1}\right) + \mathrm{I}\left(A_{i2,j2}\right) - \mathrm{I}\left(A_{i2,j2}, A_{i1,j1}\right)\end{aligned} \quad (4,\ \text{а})$$

или

$$\begin{aligned}\mathrm{I}\left(A_{i1,j1}; A_{i2,j2}\right) &= \mathrm{I}\left(A_{i1,j1}\right) - \mathrm{I}\left(A_{i1,j1} \mid A_{i2,j2}\right) = \\ &= \mathrm{I}\left(A_{i2,j2}\right) - \mathrm{I}\left(A_{i2,j2} \mid A_{i1,j1}\right) = \mathrm{I}\left(A_{i2,j2}; A_{i1,j1}\right).\end{aligned} \quad (4,\ \text{б})$$

$$\mathrm{I}\left(A_{i1,j1};A_{i2,j2}\right)=\log N+\mathbf{I}\left(A_{i1,j1};A_{i2,j2}\right),$$

где $\mathbf{I}\left(A_{i1,j1};A_{i2,j2}\right)$ *суммарная взаимная* информация между значениями $A_{i2,j2}$ и $A_{i1,j1}$.

Взаимная информация показывает, какую часть информации (в битах) о значении $A_{i1,j1}$ атрибута A_{i1} мы обнаруживаем, если становится известным значение $A_{i2,j2}$ атрибута A_{i2} и наоборот.

Пример 4. Для данных, представленных в табл. 2 имеем (log – логарифм по основанию 2):

$\mathrm{I}(A_{2,1})$ = log N – log $\mathrm{M}(A_{2,1})$ = log 16 – log 6.1 = 1.391;

$\mathrm{I}(B_1)$ = log N – log $\mathrm{M}(B_1)$ = log 16 – log 4.4 = 1.862;

$\mathrm{I}(B_1,A_{2,1})$ = log N – log $\mathrm{M}(B_1\times A_{2,1})$ = 4 – log 1.62 = 3.304;

$\mathrm{I}(B_1|\,A_{2,1}) = \mathrm{I}(B_1,\,A_{2,1}) - \mathrm{I}(A_{2,1})$ = 3.304 – 1.391 = 1.913;

$\mathrm{I}(B_1;\,A_{2,1}) = \mathrm{I}(B_1) - \mathrm{I}(B_1|\,A_{2,1})$ = 1.863 – 1.913 = – 0.050.

Суммарная энтропия нечетких данных. Обобщим понятие информации в значении $A_{i,j}$ на все j (j=1,…,m_i) значений этого атрибута A_i. В результате получим энтропию нечеткого атрибута - математическое ожидание информации. Энтропия достигает максимума, когда значения A_i равномерно распределены в домене. Энтропия зависит только от распределения случайных переменных и не зависит от их значений. Это делает энтропию универсальной мерой измерения неоднородности.

Определение 9. *Суммарная* энтропия атрибута A_i равна

$$\mathbf{H}(A_i)=\sum\nolimits_{j=1}^{m_i}\mathrm{M}\left(A_{i,j}\right)\times\mathbf{I}\left(A_{i,j}\right), \qquad (5)$$

где $\mathrm{I}(A_{i,j})$ вычисляется по соотношению (1).

Энтропия исходного атрибута равна среднему значению информации, извлекаемому при определении его значения. Энтропия результирующего атрибута равна среднему значению информации, которое следует вычислить для определения значения этого атрибута.

Определение 10. *Суммарная совместная* энтропия нечетких атрибутов A_{i2} и A_{i1} равна

$$\mathbf{H}\left(A_{i2},A_{i1}\right)=\sum_{j1=1}^{m_{i1}}\sum_{j2=1}^{m_{i2}}\begin{pmatrix}\mathrm{M}\left(A_{i1,j1}\times A_{i2,j2}\right)\times\\ \times\mathrm{I}\left(A_{i1,j1},A_{i2,j2}\right)\end{pmatrix}, \qquad (6)$$

где $\mathrm{I}\left(A_{i1,j1},A_{i2,j2}\right)$ вычисляется по соотношению (2).

Определение 11. *Суммарная условная* энтропия между нечетким атрибутом A_{i2} и значением $A_{i1,j1}$ атрибута A_{i1} равна неопределенности значения нечеткого атрибута A_{i2} в случае, если задано значение $A_{i1,j1}$ атрибута A_{i1}

$$\mathbf{H}\left(A_{i2}\mid A_{i1,j1}\right)=\sum_{j2=1}^{m_{i2}}\begin{pmatrix}\mathrm{M}\left(A_{i1,j1}\times A_{i2,j2}\right)\times\\ \times\mathrm{I}\left(A_{i2,j2}\mid A_{i1,j1}\right)\end{pmatrix}, \qquad (7)$$

где $\mathrm{I}\left(A_{i2,j2}\mid A_{i1,j1}\right)$ вычисляется по правилу (3).

Суммарная условная энтропия между нечеткими атрибутами A_{i2} и A_{i1} есть неопределенность значения атрибута A_{i2}, в случае если значение атрибута A_{i1} уже известно

$$\mathbf{H}\left(A_{i2}\mid A_{i1}\right)=\sum\nolimits_{j1=1}^{m_{i1}}\mathbf{H}\left(A_{i2}\mid A_{i1,j1}\right)\text{ бит,} \qquad (8,\text{ а})$$

или $\mathbf{H}\left(A_{i2}\mid A_{i1}\right)=\mathbf{H}\left(A_{i1},\,A_{i2}\right)-\mathbf{H}\left(A_{i1}\right)$ бит. (8, б)

Определение 12. *Суммарная взаимная информация* в нечетком атрибуте A_{i1} о нечетком атрибуте A_{i2} и обратно, отражает влияние атрибута A_{i1} на атрибут A_{i2}

$$\mathbf{I}(A_{i2};A_{i1})=\mathbf{H}(A_{i2})-\mathbf{H}(A_{i2}|A_{i1})=\mathbf{H}(A_{i1})\text{-}\mathbf{H}(A_{i1}|A_{i2}) \qquad (9,\text{ а})$$

или $\mathbf{I}(A_{i2};\,A_{i1})=\mathbf{H}(A_{i2})+\mathbf{H}(A_{i1})-\mathbf{H}(A_{i1},\,A_{i2})$ (9, б)

или $$\mathbf{I}(A_{i2};A_{i1})=\sum_{j1=1}^{m_{i1}}\sum_{j2=1}^{m_{i2}}\begin{pmatrix}\mathrm{M}\left(A_{i1,j1}\times A_{i2,j2}\right)\times\\ \times\mathrm{I}\left(A_{i2,j2};A_{i1,j1}\right)\end{pmatrix},$$

где $\mathrm{I}\left(A_{i2,j2};A_{i1,j1}\right)$ вычисляется по правилам (4, а, б).

Основное различие между *суммарной взаимной* информацией и коэффициентом корреляции это различие между энтропией и дисперсией: энтропия - мера независимой от конкретных значений атрибутов в то время, когда коэффициент корреляции определяет степень функциональной зависимости (например, линейной) между значениями атрибутов.

Пример 5. Для данных табл. 2 используя выражения (5) - (9) запишем:

$$\mathbf{H}\left(A_2\right)=\mathrm{M}\left(A_{2,1}\right)\times\mathrm{I}\left(A_{2,1}\right)+\mathrm{M}\left(A_{2,2}\right)\times\mathrm{I}\left(A_{2,2}\right)+$$
$$+\mathrm{M}\left(A_{2,3}\right)\times\mathrm{I}\left(A_{2,3}\right)=6.1\times1.391+$$
$$+5.6\times1.515+4.3\times1.896=25.119;$$

$$\mathbf{H}\left(B\right)=\mathrm{M}\left(B_1\right)\times\mathrm{I}\left(B_1\right)+\mathrm{M}\left(B_2\right)\times\mathrm{I}\left(B_2\right)+$$
$$+\mathrm{M}\left(B_3\right)\times\mathrm{I}\left(B_3\right)=4.4\text{Ч}1.862+4.4\text{Ч}$$
$$\times1.862+7.2\text{Ч}1.152=24.684;$$

$$\mathbf{H}\left(B,A_2\right)=\sum\nolimits_{jb=1}^{3}\sum\nolimits_{j=1}^{3}\mathrm{M}\left(B_{jb}\times A_{2,j}\right)\times=$$
$$\times\mathrm{I}\left(B_{jb},A_{2,j}\right)=\mathrm{M}\left(B_1\times A_{2,1}\right)\times\mathrm{I}\left(B_1,A_{2,1}\right)+$$
$$+\mathrm{M}\left(B_1\times A_{2,2}\right)\times\mathrm{I}\left(B_1,A_{2,2}\right)+\ldots+$$
$$+\mathrm{M}\left(B_3\times A_{2,3}\right)\times\mathrm{I}\left(B_3,A_{2,3}\right)=46.051;$$

$$\mathbf{H}\left(B\mid A_{2,1}\right)=\sum\nolimits_{jb=1}^{3}(B_{jb}\times A_{2,1})\times\mathrm{I}\left(B_{jb}\mid A_{2,1}\right)=$$
$$=\mathrm{M}(B_1\times A_{2,1})\times\mathrm{I}\left(B_1\mid A_{2,1}\right)+\mathrm{M}(B_2\times A_{2,1})\times$$
$$\times\mathrm{I}\left(B_2\mid A_{2,1}\right)+\mathrm{M}(B_3\times A_{2,1})\times\mathrm{I}\left(B_3\mid A_{2,1}\right)=$$
$$=1.62\times1.913+3.3\times0.886+1.18\times2.370=8.820;$$

$$\mathbf{H}\left(B\mid A_2\right)=\sum\nolimits_{j=1}^{3}\mathrm{H}\left(B\mid A_{2,j}\right)=\mathbf{H}\left(B\mid A_{2,1}\right)+$$
$$+\mathbf{H}\left(B\mid A_{2,2}\right)+\mathbf{H}\left(B\mid A_{2,3}\right)=$$
$$=8.820+8.201+3.91=20.932;$$

$$\mathbf{H}\left(B\mid A_2\right)=\mathbf{H}\left(B,A_2\right)-\mathbf{H}\left(A_2\right)=46.051-25.119=20.932;$$

$$\mathbf{I}\left(B;A_2\right)=\mathbf{H}\left(B\right)-\mathbf{H}\left(B\mid A_2\right)=24.684-20.932=3.752;$$

$$\mathbf{I}\left(B;A_2\right)=\mathbf{H}\left(B\right)+\mathbf{H}\left(A_2\right)-\mathbf{H}\left(B,A_2\right)=$$
$$=24.684+25.119-46.051=3.752.$$

Результаты вычислений представлены в табл. 3.

Таблица 3 – **Суммарные условная информация и энтропия**

	I (*Надежность* | A_i)			**H** (*Надежность* | A_i)		
	Полная	*Частичная*	*Низкая*	*Полная*	*Частичная*	*Низкая*
Исходный нечеткий атрибут A_1 (*Возраст*)						
$A_{1,1}$	1.948	1.063	1.932	3.332	3.358	3.342
$A_{1,2}$	1.328	2.522	1.226	3.068	2.547	3.040
$A_{1,3}$	3.244	3.968	0.268	1.233	0.913	0.801
H(*Надежность* | *Возраст*) = 21.632						
Исходный нечеткий атрибут A_2 (*Доход*)						
$A_{2,1}$	1.913	0.886	2.370	3.099	2.925	2.797
$A_{2,2}$	1.429	2.654	1.090	2.972	2.362	2.868
$A_{2,3}$	2.619	4.356	**0.343**	1.833	0.915	1.163
H(*Надежность* | *Доход*) = 20.932 (минимальное)						
Исходный нечеткий атрибут A_3 (*Стаж*)						
$A_{3,1}$	2.322	1.515	1.152	2.601	2.969	2.903
$A_{3,2}$	1.665	2.092	1.152	5,.61	5.104	5.391
H(*Надежность* | *Стаж*) =24.428						
Исходный нечеткий атрибут A_4 (*Залог*)						
$A_{4,1}$	2.987	2.116	0.637	2.449	3.173	2.662
$A_{4,2}$	1.408	1.712	1.653	5.041	4.964	4.993
H(*Надежность* | *Залог*) =23.283						

Суммарные взаимная информация и энтропия последовательности атрибутов. Рассмотрим последовательность исходных атрибутов $A_{i1},\ldots,A_{iq}$ и один результирующий атрибут B.

Определение 13. *Суммарная совместная* информация последовательности значений $U_q = \{A_{i1,j1},\ldots,A_{iq,jq}\}$ ($q \geq 2$) и значения B_j ($j = 1,\ldots,m_b$) вычисляется как

$$\mathbf{I}\left(B_j, U_q\right) = -\log \mathrm{M}\left(B_j \times A_{i1,j1} \times \ldots \times A_{iq,jq}\right), \quad (10)$$

Определение 14. *Суммарная условная* энтропия между атрибутами B, A_{iq} и последовательностью значений $U_{q\text{-}1} = \{A_{i1,j1},\ldots,A_{iq\text{-}1,jq\text{-}1}\}$ нечетких атрибутов $\{A_{i1},\ldots,A_{iq\text{-}1}\}$ есть неопределенность значения результирующего атрибута B, в случае, если последовательность значений $U_{q\text{-}1}$ и значение $A_{iq,jq}$ (или нечеткий атрибут A_{iq}) уже известны

$$\mathbf{H}\left(B \mid U_{q-1}, A_{iq,jq}\right) = \sum_{j=1}^{m_b} \begin{pmatrix} \mathrm{M}\left(B_j \times U_q\right) \times \\ \times\left(\mathbf{I}\left(B_j, U_q\right) - \mathbf{I}\left(U_q\right)\right) \end{pmatrix} \quad \mathbf{(11)}$$

или

$$\mathbf{H}\left(B \mid U_{q-1}, A_{iq}\right) = \sum\nolimits_{jq=1}^{m_{iq}} \mathbf{H}\left(B \mid U_{q-1}, A_{iq,jq}\right). \quad (12)$$

Определение 15. *Суммарная взаимная* информация в результирующем атрибуте B о нечетком атрибуте A_{iq} и последовательности значений $U_{q\text{-}1} = \{A_{i1,j1},\ldots,A_{iq\text{-}1,jq\text{-}1}\}$ отражает влияние атрибута A_{iq} на атрибут B, если известна последовательность нечетких значений $U_{q\text{-}1}$.

$$\mathbf{I}\left(B; U_{q-1}, A_{iq}\right) = \\ = \sum_{jq=1}^{m_{iq}} \sum_{j=1}^{m_b} \begin{pmatrix} \mathrm{M}\left(B_j \times A_{i1,j1} \times \ldots \times A_{iq,jq}\right) \times \\ \left(\mathbf{I}\left(B_j, U_{q-1}\right) + \mathbf{I}(A_{iq,jq}, U_{q-1}) - \right. \\ \left. \mathbf{I}(B_j, U_{q-1}, A_{iq,jq}) - \mathbf{I}\left(U_{q-1}\right)\right) \end{pmatrix} \quad (13, \text{а})$$

или

$$\mathbf{I}\left(B; U_{q-1}, A_{iq}\right) = \mathbf{H}\left(B \mid U_{q-1}\right) - \mathbf{H}\left(B \mid U_{q-1}, A_{iq}\right). \quad (13, \text{б})$$

Пример 6. Для данных, заданных в табл. 2, имеем B = *Надежность*, $U_{q\text{-}1} = \{A_{2,1}\}$ = {*Высокий доход*}. Используя выражения (11, 12 и 13, б) получим:

$$\mathbf{H}\left(B \mid A_{2,1}, A_{1,1}\right) = \\ = \sum_{jb=1}^{m_b} \begin{pmatrix} \mathrm{M}\left(B_{jb} \times A_{2,1} \times A_{1,1}\right) \times \\ \times\left(\mathrm{I}\left(B_{jb}, A_{2,1}, A_{1,1}\right) - \mathrm{I}\left(A_{2,1}, A_{1,1}\right)\right) \end{pmatrix} = \\ = 1.935 + 1.564 + 1.227 = 4.725;$$

$$\mathbf{H}\left(B \mid A_{2,1}, A_1\right) = \mathbf{H}\left(B \mid A_{2,1}, A_{1,1}\right) + \mathbf{H}\left(B \mid A_{2,1}, A_{1,2}\right) + \\ + \mathbf{H}\left(B \mid A_{2,1}, A_{1,3}\right) = 4.725 + 1.927 + 1.000 = 7.652;$$

$$\mathbf{I}\left(B; A_{2,1}, A_1\right) = \mathbf{H}\left(B \mid A_{2,1}\right) - \mathbf{H}\left(B \mid A_{2,1}, A_1\right) = \\ = 8.820 - 7.652 = 1.168.$$

Результаты вычислений приведены в табл. 4.

Таблица 4 – **Суммарные условные информация и энтропия последовательности атрибутов, при известном значении *Доход = Высокий***

	I (*Надежность* | *Доход* = *Высокий*, A_i)			**H** (*Надежность* | *Доход* = *Высокий*, A_i)		
	Полная	*Частичная*	*Низкая*	*Полная*	*Частичная*	*Низкая*
$A_{1,1}$	1.931	0.636	3.408	1.935	1.564	1.227
$A_{1,2}$	1.411	0.990	3.055	0.732	0.688	0.507
$A_{1,3}$	3.184	2.614	0.461	0.315	0.384	0.301
$A_{3,1}$	2.547	0.635	2.435	1.268	1.190	1.310
$A_{3,2}$	1.507	1.162	2.313	1.691	1.656	1.485
$A_{4,1}$	2.674	0.901	1.701	0.880	1.013	1.099
$A_{4,2}$	1.632	0.879	2.905	2.106	1.912	1.550

Построение деревьев нечетких решений

Поясним использование суммарных информационных оценок при построении деревьев нечетких решений. Заметим, что для построения дерева нечетких решений используется только одна из этих оценок – *суммарная условная* энтропия. *Суммарная взаимная* информация используется при ограничении размера дерева (формирования листьев).

Существует два аспекта при построении нечеткого дерева решений. Один из них – выбор очередного атрибута в качестве узла дерева - атрибута имеющего минимальную *суммарную условную* энтропию. Другой аспект – ограничение роста дерева, отсечение непродуктивных ветвей. При этом используются два типа ограничений (порога) β и α. Узлы трансформируются в листья если:

(а) частота возникновения ветви $U_q = \{A_{i1,j1},\ldots,A_{iq,jq}\}$ меньше или равно, чем заданное ограничение α: $\mathbf{I}(U_q) \geq -\log \alpha \times N$.

(б) относительная частота одного значения больше или равно, чем заданное ограничение β :

$$\min \mathbf{I}(B_j \mid U_q) \leq -\log \beta \quad \text{for } \forall j = 1,\ldots m_b;$$

Уменьшение β и увеличение α приводит к построению дерева меньшей размерности. Однако точность оценки при этом также снижается. Влияние изменения параметров β и α на результат представлено в работе [18].

Алгоритм построения нечеткого дерева решений выглядит следующим образом [19].

Исходные данные: Наборы исходных и результирующего атрибутов (см. табл. 2).

$Attr = \{A_1,\ldots,A_n\} \in \boldsymbol{I}$; $B \in \boldsymbol{O}$; α, β; $q=0$; $U_q = \varnothing$.

Выходные данные: Дерево нечетких решений.

$Tree$ = buildTree $(U_q$, Attr$)$.

{ **1**. Вычисление *суммарной* энтропии (8, 11)

2. Выбор атрибута с минимальной энтропией

$num = \operatorname{argmin} \mathbf{H}(B|U_q \cup A_i)$ for $\forall$ $A_i \in Attr$,

3. Назначение текущего узла

Tree $\leftarrow$ node (A_{num}); $Attr = Attr \backslash A_{num}$;

4. Выбор листьев и продолжение

```
for (∀ A_num,j ), j=1,…,m_num;
    { q++;              U_q = U_q-1 ∪ A_num,j ;
      if   (A_num,j is leaf) Tree ← leaf (A_num,j)
      else Рекурсивное построение поддерева:
           Tree = buildTree (U_q, Attr)  }
}
```

Поясним использование алгоритма на примере.

Пример 7. *Исходные данные:* Наборы исходных и результирующего атрибутов, содержащие 4 исходных атрибута $\boldsymbol{I}$ = {*Возраст*, *Доход*, *Стаж*, *Залог*} (n=4) и одного результирующего атрибута *Надежность* (табл. 2). Выберем ограничения равными α=0.25 и β=0.75.

Выходные данные: Дерево нечетких решений

Шаг 1. Вычислим *суммарную условную* энтропию между исходными и результирующим атрибутами в соответствии с выражением (8), табл. 3.

Шаг 2. Выберем атрибут A_i с минимальной энтропией:

$num = \operatorname{argmin} \mathbf{H}(\text{\textit{Надежность}}\,|A_i)$ for $\forall$ $A_i \in \boldsymbol{I}$.

$num = \operatorname{argmin} \{21.632;\ 20.932;\ 24.428;\ 23.283\} = 2$.

Шаг 3. В качестве текущего узла выберем атрибут A_2 = *Доход*.

Шаг 4. Проверим, является ли данный узел листом. $\min\limits_{\text{for } \forall j2} \mathbf{I}(A_{2,j2}) \le -\log(0.25\times 16)$:

$min\ \mathbf{I}(A_{2,1}) = \mathbf{I}(\text{\textit{Высокий}}) = -2.609$ бит (нет)

$min\ \mathbf{I}(A_{2,2}) = \mathbf{I}(\text{\textit{Средний}}) = -2.485$ бит (нет)

$min\ \mathbf{I}(A_{2,1}) = \mathbf{I}(\text{\textit{Низкий}}) = -2.104$ бит (нет)

$$\min_{\text{for } \forall j} \mathbf{I}(B_j|A_{2,j2}) \le -\log 0.75 = 0.415 \text{ бит}$$

$min\ \mathbf{I}(B_2\,|\,A_{2,1}) = \mathbf{I}(\text{\textit{Полная}}\,|\,\text{\textit{Высокий}}) = 0.886$ (нет)

$min\ \mathbf{I}(B_3\,|\,A_{2,2}) = \mathbf{I}(\text{\textit{Низкая}}\,|\,\text{\textit{Средний}}) = 1.090$ (нет)

$min\ \mathbf{I}(B_3\,|\,A_{2,3}) = \mathbf{I}(\text{\textit{Низкая}}\,|\,\text{\textit{Низкий}}) = \mathbf{0.343}$ (**лист**)

Строим первый уровень дерева нечетких решений (рис. 3).

Шаг 5. Повторить *Шаг* 1 – *Шаг* 4 для двух первых ветвей и трех оставшихся исходных атрибутов {*Возраст*, *Стаж* и *Залог*}. Результаты вычислений для первой ветви приведены в табл. 4, а само дерево нечетких решений представлено на рис. 4.

Конец алгоритма.

Поясним первый уровень полученного дерева решений с информационной точки зрения. Изначально, для оценки надежности возврата кредита требуется получить **H**(*Надежность*) = 24.684 бит информации.

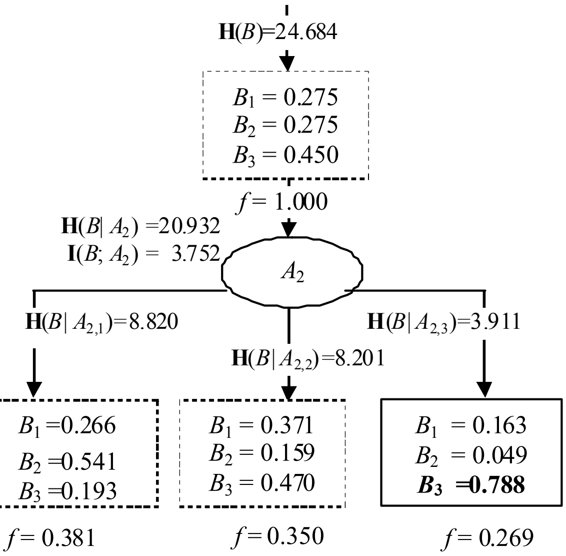

Рис. 3. Первый уровень дерева нечетких решений с одним узлом и одним окончательным листом

В качестве первого оцениваемого критерия выберем A_2 = *Доход*, определение значения которого вносит максимальное количество информации

I(*Надежность*; *Доход*) = 3.752 бита.

В результате, нам в дальнейшем требуется извлечь не более **H**(*Надежность* | *Доход*) = 20.932 бит информации.

При этом, неопределенность информации распределена следующим образом:

H(*Надежность* | *Доход = Высокий*) = 8.820 бит;

H(*Надежность* | *Доход = Средний*) = 8.201 бит;

H(*Надежность* | *Доход = Низкий*) = 3.911 бит.

Если оценка показателя *Доход* нового заемщика *равна Низкий*, то с уверенностью 78.8% в качестве итоговой оценки надежности следует принять значение *Низкая* надежность. Это значение является листом дерева нечетких решений, поскольку в данном случае относительная частота значения *Низкая* = 78.8% больше заданного ограничения β=0.75 или 75%. Для остальных вариантов *Доход = Высокий* или *Доход = Средний* требуется построить второй уровень дерева нечетких решений (рис. 4).

Процесс работы разбит на 2 этапа: (*а*) обучение и настройка СППР и (*б*) использование СППР.

Первый этап выполняется однократно. На этом этапе осуществляется построение нечеткого дерева решений (синтез продукционных правил) и как результат - происходит обучение СППР. Это обучение предполагает проверку выборки предоставленных кредитов и накопление информации о результатах надежности их возврата. Необходимым условием здесь является получение репрезентативной выборки.

Второй этап предполагает работу СППР. Изначально проверяется параметр, ответ на который позволяет извлечь максимальную информацию о надежности возврата кредита. В зависимости от ответа проверяется следующий параметр, также извлекающий максимальную информацию о надежности возврата. Процесс задания вопросов происходит до тех пор, пока уверенность в оценке надежности не достигает заданного ограничения α или β.

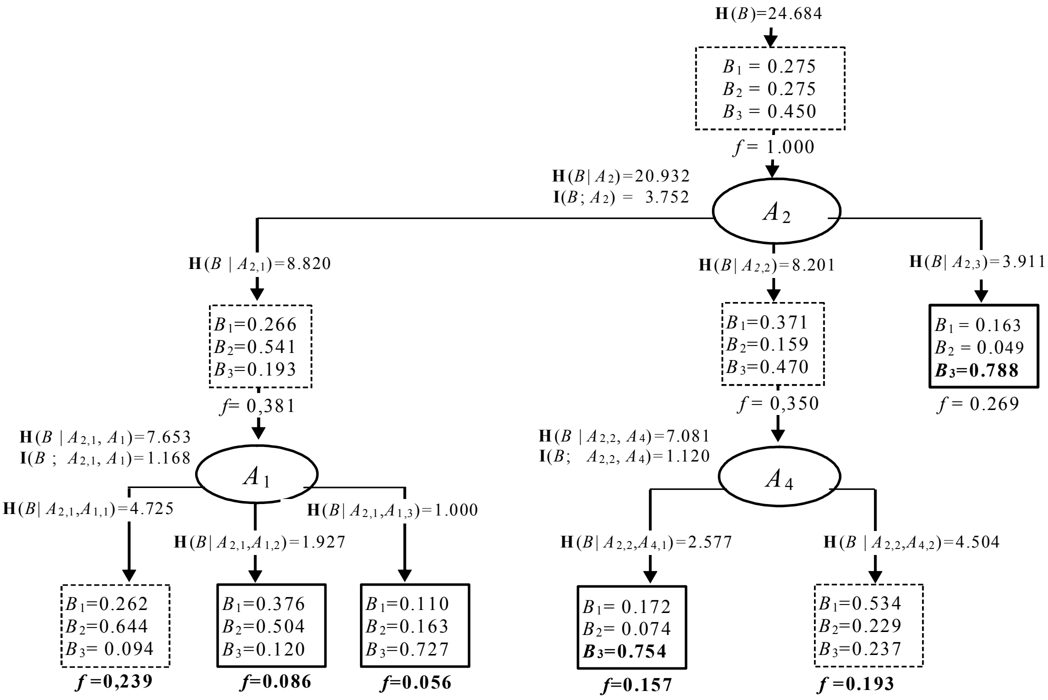

Рис 4. Дерево нечетких решений, описывающее стратегию принятия решения о надежности возврата кредита и построенное на основе данных табл. 2 с ограничениями α=0.25 и β=0.75

Заключение и направления будущих исследований

В работе предложен математический аппарат построения СППР, предусматривающих анализ предыдущих результатов проверки и выработку стратегии тестирования. В основу математического аппарата положены суммарные информационные оценки нечетких данных и алгоритм построения на их основе деревьев нечетких решений Наличие пороговых ограничений β и α позволяет строить деревья заданной размерности и заданной точности. Использование информационных оценок позволяет наблюдать процесс увеличения информации и уменьшения энтропии по мере получения результатов проверок показателей, описывающих объект или процесс. Это соответствует закону сохранения информации. В дальнейших работах авторы планируют сопоставить стратегии принятия решений на основе деревьев нечетких решений со стратегиями, полученными на основе применения нейронных сетей, в том числе и нейронных сетей на основе конволюции.

Благодарность

Работа была поддержана грантом *Slovak Research and Development Agency* регистрационный номер APVV-18-0027 «*New methods development for reliability analysis of complex system*».

References

1. Enoki, H., Takami, K., Kobayashi, Y. and Ohta, T. (1994), “Switching function conceptual model for telecommunication service specification design”, *Proc. of the IEEE Global Telecommunications Conf.*, vol. 2, San Francisco, USA, pp. 818-822.
2. Levashenko, V., Zaitseva, E., Kovalík, Š. (2013), *Projektovanie systémov pre podporu rozhodovania na základe neurčitých dát*, EDIS Publ., Zilina, Slovakia, ISBN 978-80-554-0680-0.
3. Quinlan, J.R. (1990), “Decision Tree and Decision Making”. *IEEE Trans. on Syst., Man and Cyb.*, vol. 20, no. 2, pp. 339-346.
4. Hullermeier, E. and Vanderlooy, S. (2009), “Why Fuzzy Decision Trees are Good Rankers”, *IEEE Transactions on Fuzzy Systems, v*ol. 17, no. 6, pp. 1233-1244.
5. Yong, S., Lin, M., Robinson, W. and Fidge, C. (2010), “Using decision trees in economizer repair decision making”, *Proc. of the Int. Conf. on Prognostics and Health Management* (PHM), Portland, USA, pp. 1-6.
6. Lee, H-M., Chen, C.-M., Chen, J.-M. and Jou, Y.-L. (2001), “An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy”, *IEEE Transactions on Systems, Man and Cybernetics*, Part B: *Cybernetics*, vol. 31, no. 3, pp. 426-432.
7. Garcia, S., Luengo, J., Saez, J., Lopez, V. and Herrera, F. (2013), “A Survey of Discretization Techniques: Taxonomy and Empirical Analysis in Supervised Learning”, *IEEE Trans. on Knowledge and Data Engineering,* vol. 24, no. 5, pp. 734-750.
8. Yuan, Y. and Shaw, M.J. (1995), “Induction of Fuzzy Decision Trees”, *Fuzzy Sets and Systems*, vol. 69, pp. 125-139.
9. Wang, X., Chen, B., Qian, G. and Ye, F. (2000), “On the Optimization of Fuzzy Decision Trees”, *Fuzzy Sets and Systems*, vol. 112, pp. 117–125.
10. Levashenko, V., Zaitseva, E. and Puuronen, S. (2007), “Fuzzy classifier based on fuzzy decision tree”, *Proc. of the IEEE Int. Conf. on Computer as a Tool* (EUROCON), Warsaw, Poland, pp. 823– 827.
11. Gasir, F., Crockett, K., and Bandar, Z. (2012), “Inducing fuzzy regression tree forests using artificial immune systems”, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 20, pp. 133–157.
12. Zhai, J., Wang, X., Zhang, S. and Hou, S. (2018), “Rough fuzzy decision tree”, *Information Sciences*, vol. 465, pp.425–438.
13. Ahmadi, H., Gholamzadeh, M., Shahmoradi, L., Nilashi, M. and Rashvand, P. (2018), “Diseases diagnosis using fuzzy logic methods: A systematic and meta-analysis review”, *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 145–172.

14. Olaru, C. and Whenkel, L. (2003), “A complete fuzzy decision tree technique”, *Fuzzy Sets and Systems*, vol.138, pp.221–254
15. Rabcan, J., Levashenko, V., Zaitseva, E., Kvassay, M. and Subbotin, S. (2019), “Application of fuzzy decision tree for signal classification", *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5425-5434, doi: 10.1109/TII.2019.2904845.
16. Levashenko, V., Zaitseva, E., Kvassay, M. and Deserno T. (2016), “Reliability estimation of healthcare systems using fuzzy decision trees”, *Proc. of the Fed. Conf. on Comp. Science and Inf. Systems* (FedCSIS), Gdansk, Poland, 2016, pp. 331–340.
17. Kurbatsky, A.N. and Cheushev, V.A. (1999), *Informational method of analysis and optimization in decision support systems*. Ed. Institute of Technical Cybernetics, Minsk, Belarus.
18. Zaitseva, E. and Levashenko, V. (2016), “Construction of a reliability structure function based on uncertain data”, *IEEE Transaction on Reliability*, vol. 65, no. 4, pp. 1710–1723.
19. Zaitseva, E., Levashenko, V., Rabcan, J. and Krsak, E. (2020), “Application of the Structure Function in the Evaluation of the Human Factor in Healthcare”, *Symmetry Basel*, vol.12, no. 93; doi:10.3390/sym12010093.

Received (Надійшла) 02.10.2020
Accepted for publication (Прийнята до друку) 04.11.2020

Відомості про авторів / About the Authors

Левашенко Віталій Григорович – професор, завідувач кафедри інформатики, факультет управлінських наук та інформатики, Жилінський університет, Жиліна, Словаччина;
Vitaly Levashenko – Professor, Head of the Department of Informatics, Faculty of Management Science and Informatics, University of Žilina, Žilina, Slovakia;
e-mail: vitaly.levashenko@fri.uniza.sk; ORCID ID: https://orcid.org/0000-0003-1932-3603

Ляшенко Олексій Сергійович – кандидат технічних наук, доцент, декан факультету комп'ютерної інженерії та управління, Харківський національний університет радіоелектроніки, Харків, Україна;
Oleksii Liashenko – Candidate of Technical Sciences, Associate Professor, Dean of Faculty of Computer Engineering and Control, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: oleksii.liashenko@nure.ua; ORCID ID: http://orcid.org/ 0000-0002-0146-3934.

Кучук Георгій Анатолійович – доктор технічних наук, професор, професор кафедри обчислювальної техніки та програмування, Національний технічний університет “Харківський політехнічний інститут”, Харків, Україна;
Heorhii Kuchuk – Doctor of Technical Sciences, Professor, Professor of Computer Science and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: kuchuk56@ukr.net; ORCID ID: http://orcid.org/0000-0002-2862-438X

Побудова системи підтримки прийняття рішень на основі нечітких даних

В. Г. Левашенко, О. С. Ляшенко, Г. А. Кучук

Анотація. Розробка інструментарію оцінки прийнятих рішень є актуальним і потрібним завданням на сучасному етапі розвитку інформаційних технологій. Таким інструментарієм є, наприклад, системи підтримки прийняття рішень (СППР). У роботі пропонується математичний апарат побудови СППР. Побудова СППР передбачає аналіз наявних результатів спостережень або вимірювань і вироблення стратегії перевірок вихідних параметрів у вигляді дерева нечітких рішень або продукційних правил. Основу запропонованого апарата становлять сумарні інформаційні оцінки (інформація і ентропія) для нечітких наборів даних. Використання нечітких даних найбільш повно відповідає людській природі, оскільки на практиці люди часто застосовують суб'єктивні відчуття і апріорні знання, ніж точні імовірнісні критерії. Тому, використовуючи нечітку логіку і розглядаючи ступінь можливості як нечітку міру, експерти мають можливість описувати реальні дані з достатньою точністю. Досліджено взаємозв'язок запропонованих сумарних інформаційних оцінок. В роботі наведені приклади, що демонструють використання запропонованого математичного апарату для практичних задач. У подальшій роботі автори планують привести результати експериментальних досліджень запропонованого підходу і його зіставлення з іншими відомими методами і алгоритмами. Зазначене зіставлення представляється для широкого різноманіття формалізованих даних, що зберігаються в відомому репозиторії UCI Machine Learning Repository. У якості зіставлених методів і алгоритмів планується вибрати інші алгоритми побудови дерев нечітких рішень, алгоритми байєсівської класифікації, побудови дерев рішень C4.5, CART і метод найближчих сусідів.

Ключові слова: системи підтримки прийняття рішень; нечітка логіка; дерева нечітких рішень.

Building Decision Support Systems based on Fuzzy Data

Vitaly Levashenko, Oleksii Liashenko, Heorhii Kuchuk

Abstract. The implementation of tools for assessing decisions making is an urgent and demanded task at the current stage of information technology development. Decision-making support systems (DMSS) are perspective tools for it. The paper proposes a mathematical tool for construction a DMSS. DMSS construction involves the analysis of the available observation or measurement results and the development of a strategy for checking the initial parameters in the form of a fuzzy decision tree or production rules. The proposed tool is based on cumulative information estimates (information and entropy) for fuzzy data sets. The use of fuzzy data is a most fully consistent with human nature. It can be explained, that in practice people often use subjective sensations and a priori knowledge than precise probabilistic criteria. Therefore, using fuzzy logic and considering the degree of possibility as a fuzzy measure, experts are able to describe real data with sufficient accuracy. The relationship of the proposed total information estimates is investigated in the paper. The paper provides examples which demonstrating the practical application of the proposed mathematical tool. The authors plan to present the results of experimental investigations of the proposed approach and its comparison with other known methods and algorithms in a next work. These results will be obtained on a wide range of formalized data stored in the well-known UCI Machine Learning Repository. As compared methods and algorithms, we are going to choose algorithms for constructing fuzzy decision trees based on Luca de Termini entropy, Naive-Bayesian classification, algorithms for decision trees induction - C4.5, CART and the method of nearest neighbours.

Keywords: decision making support systems; fuzzy logic; fuzzy decision trees.