

С. Г. Лелеков, канд. физ.-мат. наук, доц., e-mail: p1859@yandex.ru,
 Севастопольский государственный университет,
 А. М. Лях, канд. биол. наук, ст. науч. сотр., e-mail: antonlyakh@yahoo.com,
 Институт биологии южных морей, г. Севастополь

Нечеткий вывод в адаптивных таксономических экспертных системах

Семантическая неоднозначность естественного языка вызывает проблему представления и манипулирования знаниями в таксономических экспертных системах. Для представления признаков определяемых объектов предлагается использовать лингвистические переменные с нечеткими значениями. Рассматриваются способы организации нечеткого вывода в экспертных системах.

Ключевые слова: экспертные системы, лингвистические переменные, нечеткий вывод, биологическая таксономия, компьютерные определители

Введение

Любая идентификация (определение) организмов, в том числе компьютерная, основывается на данных о признаках объектов и их таксономическом статусе. Поэтому компьютерные таксономические базы данных (БД) и атласы организмов — это наиболее распространенный и активно используемый сетевой ресурс. Они содержат списки таксонов определенной группы, например, *algaebase.org* — таксономическая БД о водорослях, *marinespecies.org* — таксономическая БД морских организмов, *tolweb.org* — таксономическая БД всех организмов, *fishbase.org* — описание рыб Мирового океана и др. Эти сетевые ресурсы не являются экспертными системами, так как не содержат правил определения и не предоставляют интерфейс для идентификации. Они формируют базу таксономических знаний, которая может быть использована для создания компьютерных определителей.

Современные компьютерные определители применяют полиномические одно- или многовходовые ключи, заданные таблицами признаков. Процесс определения является пошаговым. Признаки выбираются алгоритмами системы на основе их диагностической ценности, которая рассчитывается по таблице признаков. Процедура идентификации продолжается до опознавания таксона или исчерпания признаков. Результатом является перечень таксонов, у которых признаки частично или полностью совпадают с введенными пользователем значениями. Интуитивно понятный механизм взаимодействия с пользователем позволяет корректно идентифицировать организмы и не специалисту.

Такой подход реализован в системах "Lucid" (*lucidcentral.org*), "Delta-IntKey" (*delta-intkey.com*), "Taxex"[1—4].

Однако существующие таксономические экспертные системы пока не находят широкого применения. Одной из причин этого является семан-

тическая неоднозначность естественного языка, используемого экспертами.

При составлении диагнозов таксонов эксперты используют своеобразную терминологию, которая вплетается в живой разговорный язык. Диагнозы часто содержат размытые формулировки, понятные автору, но не понятные стороннему читателю, например, признак "Форма головы" со значениями: "Большая массивная", "Небольшая", "Маленькая". Пользователь вынужден принимать терминологию эксперта, хотя его представление о сущности того или иного термина может отличаться от смысла, который вложил в него эксперт. Проблема усугубляется тем, что общение пользователя и эксперта проходит опосредованно через экспертную систему.

Частичным решением этой проблемы может стать использование адаптивных таксономических экспертных систем [5], в которые закладываются механизмы учета семантической неоднозначности естественного языка. Математической основой функционирования таких систем является теория нечетких множеств. Адаптивность, в этом контексте, следует понимать как приспособляемость системы к нечеткой интерпретации терминов пользователем и экспертом в процессе таксономической идентификации объекта.

Постановка задачи

В заданной предметной области имеется конечное множество визуально распознаваемых объектов. Все объекты описаны множеством признаков. Описания выполнены на естественном языке и, как следствие, содержат нечеткие, расплывчатые формулировки.

Требуется разработать автоматизированную диагностическую систему, способную в диалоговом режиме работы с пользователем идентифицировать предъявляемый ей объект путем последовательного выяснения значений признаков. Такие системы в

биологии получили название *адаптивных таксономических экспертных систем* (АТЭС).

Рассматривается решение задачи сочетания неоднозначности описаний биологических объектов, их толкования и способов организации логического вывода в рамках компьютерных адаптивных таксономических экспертных систем.

Эта задача включает две подзадачи:

- описание определяемых объектов в базе признаков;
- использование описаний в процессе логического вывода.

В статье предложено решение в основном второй подзадачи. Решение первой задачи более подробно рассмотрено в работе [6].

Описание определяемых объектов в базе признаков

База знаний АТЭС включает: описания признаков, описания таксонов и набор правил вывода заключительного диагноза.

Простейшим вариантом представления правил вывода является форма:

ЕСЛИ " α есть β ", ТО " γ есть δ ".

При этом " α есть β " и " γ есть δ " являются нечеткими высказываниями [7].

В рассматриваемом классе экспертных систем (ЭС) α является таксономическим признаком, β — его значением, γ — неизвестным определяемым объектом, δ — подмножеством известных системе таксонов. Наибольшие трудности при переносе знаний эксперта в базу знаний ЭС связаны с формированием машинного представления α и β .

Представление признаков реализуется в двух формах:

- внешней, в виде веб-страницы с текстом и рисунками, отражающими значения признака;
- внутренней, в которой учитывается семантическая неопределенность естественного языка.

Для внутреннего представления предлагается использовать аппарат нечетких множеств и понятие лингвистической переменной [4].

Лингвистическая переменная определяется как

$$L = (\chi, T, U, M),$$

где χ — название лингвистической переменной; T — совокупность лингвистических значений L , выраженных нечеткими переменными; U — универсальное множество (область определения T); M — семантические правила, которые каждому лингвистическому значению $x \in T$ ставят в соответствие нечеткое подмножество $M(x)$, которое может быть задано функцией принадлежности $\mu: U \rightarrow [0, 1]$.

Рассмотрим особенности представления таксономических признаков.

Все признаки можно разделить:

- на группу A , для которых на множестве U можно задать линейный порядок;

- на группу B , для которых на множестве U линейный порядок задать нельзя.

В группу A войдут:

- количественные признаки, например, признак "Число щетинок на первой паре торакокод" со значениями 1, 2, 3, 4, или признак "Длина тела";
- признаки, в которых неявно используется количественная шкала, например, признак "Длина антенн I" со значениями "Длинные, больше длины раковины", "Почти равные длине раковины", "Короткие, меньше длины раковины" или признак "Наличие выростов на голове" со значениями "Без выростов", "С незначительными выростами", "С выростами".

Последний пример иллюстрирует необходимость учета слов-модификаторов.

В русском языке модификаторы представлены в основном количественными наречиями, которые можно разбить на три группы [8].

В первую группу входят те, которые "размывают" нечеткие значения лингвистических переменных, например, *более-менее горячий, около пяти, где-то 12* и т. п.

Назовем эту группу группой типа "Равно" и включим в нее следующие модификаторы: *приблизительно, около, где-то, почти, примерно, вероятно, кажется, более или менее, скорее всего, по-моему*.

Во вторую группу типа "Больше — Меньше" войдут модификаторы, которые изменяют значение переменной в одну из сторон: *немного, чуть, значительно, очень, слегка, несколько, много, не более, не менее, сильно, слишком, едва*. Например: *немного горячий, почти холодный, очень горячий* и т. п.

В третью группу войдут те, которые акцентируют, уменьшают неопределенность нечетких значений лингвистической переменной, например, *конечно это...*. Варианты: *определенно, точно* и т. п.

Использование модификаторов приводит к необходимости расширения множества U в соответствии с принципом обобщения для нечетких множеств [7, 9].

Приведенные примеры предполагают линейный порядок элементов на множестве определения U нечетких значений лингвистических переменных заданием некоторой количественной меры. Примером может служить температура, радиус кривизны, скорость, количество и т. п.

Рассмотрим вторую группу признаков (B), где линейный порядок на множестве U либо не существует, либо его задать нельзя.

Причиной несущественности линейного порядка может быть наличие всего двух взаимно ортогональных значений лингвистической переменной: *да/нет, истина/ложь, он/не он* и т. п. Например, признак "Зубцы на кромке раковины моллюска" со значениями: "Есть", "Отсутствуют".

Линейный порядок на множестве U нельзя задать, когда одно или несколько значений выбира-

ется из конечного неупорядоченного множества. Например, значения признака "Форма раковины" — это неупорядоченное множество: "Треугольная", "Треугольно-эллиптическая", "Эллиптическая", "Круглая". То же самое справедливо для признака "Цвет пигментных пятен": "Желтый", "Бурый", "Сиреневый". В таких случаях формирование функций принадлежности может осуществляться, например, с помощью метода попарных сравнений [9].

Результатом описания признаков лингвистическими переменными является таблица, структура которой имеет классический вид: столбцы — признаки, строки — объекты, в ячейках — нечеткие значения признаков для соответствующих объектов.

Использование описаний в процессе логического вывода

Пусть для каждого объекта базы знаний имеется некоторая результирующая оценка — степень соответствия описания объекта в пространстве признаков данным, полученным от пользователя. В начале работы все результирующие оценки равны, например, единице. Другими словами, все объекты из базы знаний системы претендуют быть результатом идентификации.

Тогда последовательность действий экспертной системы при диагностировании объекта можно описать следующим алгоритмом.

1. Система выбирает некоторый признак, представленный лингвистической переменной, и демонстрирует пользователю его внешнее описание.

2. Пользователь выбирает одно или несколько значений признака; при этом он может использовать слова-модификаторы.

3. Система формирует нечеткое значение лингвистической переменной, соответствующей признаку.

4. Для каждого объекта, представленного в базе данных, система вычисляет оценку степени соответствия построенной лингвистической переменной значениям признаков объектов.

5. Полученная оценка учитывается в общей результирующей оценке.

6. Система повторяет п. 1 или завершает процесс по достижению некоторого критерия для результирующей оценки.

Пример. Рассмотрим работу алгоритма на примере.

1. Пусть в базе данных представлена информация о четырех объектах, и вначале результирующая оценка у всех объектов равна 1:

$$Res_1 = 1; Res_2 = 1; Res_3 = 1; Res_4 = 1.$$

2. Система спрашивает пользователя о "Форме раковины моллюска".

Предположим, формальное представление этого признака в виде лингвистической переменной следующее:

$$L_1 = (\text{"Форма раковины"}, \{\text{"треугольная"}, \text{"треугольно-эллиптическая"}, \text{"эллиптическая"}, \text{"круглая"}\}, \{\text{"Объект 1"}, \text{"Объект 2"}, \text{"Объект 3"}, \text{"Объект 4"}\}),$$

$$M = \{\mu(\text{"треугольная"}), \mu(\text{"треугольно-эллиптическая"}), \mu(\text{"эллиптическая"}), \mu(\text{"круглая"})\}.$$

Пользователь отвечает, что у его объекта форма раковины — треугольная, но он не совсем в этом уверен.

Рассмотрим действия системы.

Пусть функции принадлежности множества M признака L_1 имеют следующий вид:

- $\mu(\text{"треугольная"}) = \{0,9, 0,3, 0,5, 0,4\}$;
- $\mu(\text{"треугольно-эллиптическая"}) = \{0,5, 0,8, 0,5, 0,7\}$;
- $\mu(\text{"эллиптическая"}) = \{0,1, 0,5, 0,8, 0,4\}$;
- $\mu(\text{"круглая"}) = \{0,2, 0,3, 0,5, 0,9\}$.

Так как выбрано значение признака "треугольная", то результирующую оценку для следующего шага определения Res_i^{j+1} можно вычислить по формуле [9]

$$Res_i^{j+1} = \min(Res_i^j, \mu_i(z)), \quad (1)$$

где z — значение признака.

Так как пользователь не совсем уверен в точности определения, необходима поправка

$$\mu_i^*(z) = \begin{cases} \mu_i(z)k, & \text{если } \mu_i(z)k \leq 1; \\ 1, & \text{если } \mu_i(z)k > 1. \end{cases}$$

Возможные значения коэффициента k приведены в таблице.

Значение μ_i признака	Сильно сомневаюсь	Не совсем уверен	Согласен	Почти уверен	Абсолютно уверен
Больше или равно 0,7	k^2	k	1	$1/k$	$1/k^2$
Меньше 0,7	$1/k^2$	$1/k$	1	k	k^2

Значение k выбирается из интервала $[0,90...0,95]$.

Выполняем коррекцию результирующей оценки. При $k = 0,95$ получим $Res_1 = 0,86$; $Res_2 = 0,32$; $Res_3 = 0,53$; $Res_4 = 0,42$.

3. Пользователю предъявляется признак "Зубцы на кромке раковины".

Формальное представление этого признака в виде лингвистической переменной, например, следующее:

$$L_2 = (\text{"Зубцы на кромке раковины"}, \{\text{"есть"}, \text{"нет"}\}, \{1, 2, 3\}),$$

$$M = \{\mu(\text{"есть"}), \mu(\text{"нет"})\}.$$

Предположим, что функции принадлежности множества \mathbf{M} признака L_2 имеют следующий вид:

$$\begin{aligned}\mu(\text{"есть"}) &= \{1,0, 0,505, 0,01\}; \\ \mu(\text{"нет"}) &= \{0,01, 0,505, 1,0\}.\end{aligned}$$

Пользователь отвечает, что у его объекта зубцы на кромке раковины, кажется, есть.

Так как использовано слово — модификатор "кажется", применим операцию размывания [8] к функции $\mu(\text{"есть"})$. Смысл этой операции — сгладить отличия в значениях функции принадлежности. Получим $\mu_{ans}(\text{"есть"}) = \{1,0, 0,93, 0,36\}$.

Пусть в таблице признаков признак L_2 на множестве объектов имеет следующие значения:

$$\begin{aligned}L_2(\text{Объект1}) &= \text{"есть"}; \\ L_2(\text{Объект2}) &= \text{"есть"}; \\ L_2(\text{Объект3}) &= \text{"нет"}; \\ L_2(\text{Объект4}) &= \text{"нет"}.\end{aligned}$$

Признак L_2 задается нечеткими переменными "есть", "нет". Для вычисления результирующей оценки Res_i^{j+1} найдем пересечение значений функций принадлежности этих переменных со значениями функции $\mu_{ans}(\text{"есть"})$. Поэлементное пересечение вычисляется нахождением минимальных значений из соответствующих пар по всем элементам множества \mathbf{U} .

Тогда

$$\begin{aligned}\mathbf{d}_1 &= \mu_{ans}(\text{"есть"}) \cap \mu(\text{"есть"}) = \{1,0, 0,505, 0,01\}; \\ \mathbf{d}_2 &= \mu_{ans}(\text{"есть"}) \cap \mu(\text{"нет"}) = \{0,01, 0,505, 0,36\}.\end{aligned}$$

Для пересчета результирующей оценки выбираются максимальные значения векторов \mathbf{d}_1 и \mathbf{d}_2 . Это 1,0 и 0,505 соответственно. Тогда в соответствии с формулой (1)

$$Res_i^{j+1} = \min(Res_i^j, d_{\max}) \quad (2)$$

и $Res_1 = 0,86$; $Res_2 = 0,32$; $Res_3 = 0,505$; $Res_4 = 0,42$.

4. Система спрашивает пользователя о "Длине антенн I".

Формальное представление этого признака в виде лингвистической переменной, например, следующее:

$$\begin{aligned}L_3 &= (\text{"Длина антенн I"}, \\ \{x_1 &= \text{"длинные, больше длины раковины"}, \\ x_2 &= \text{"почти равные длине раковины"}, \\ x_3 &= \text{"короткие, меньше длины раковины"}\}, \\ \mathbf{U} &= \{1, 2, 3, 4, 5, 6, 7, 8, 9\}, \\ \mathbf{M} &= \{\mu(x_1), \mu(x_2), \mu(x_3)\}.\end{aligned}$$

Функции принадлежности признака L_3 имеют следующий вид:

$$\begin{aligned}\mu(x_1) &= \\ &= \{0,01, 0,505, 1,00, 0,750, 0,50, 0,255, 0,01, 0,010, 0,01\}; \\ \mu(x_2) &= \\ &= \{0,01, 0,255, 0,50, 0,750, 1,00, 0,750, 0,50, 0,255, 0,01\}; \\ \mu(x_3) &= \\ &= \{0,01, 0,010, 0,01, 0,255, 0,50, 0,750, 1,00, 0,505, 0,01\}.\end{aligned}$$

Пользователь отвечает, что у объекта длина антенн I чуть больше длины раковины.

Этот ответ можно интерпретировать как среднее между значениями "Антенны I длинные, больше длины раковины" и "Антенны I почти равны длине раковины". Применяя операцию размывания к $\mu(x_1)$ и $\mu(x_2)$ и усредняя значения, получим:

$$\mu_{ans} = \{0,01, 0,2, 0,71, 1,0, 0,71, 0,3, 0,04, 0,005, 0,005\}.$$

Пусть в таблице признаков признак L_3 на множестве объектов имеет следующие значения:

$$\begin{aligned}L_3(\text{Объект1}) &= x_1; \\ L_3(\text{Объект2}) &= x_1 \text{ или } x_2; \\ L_3(\text{Объект3}) &= x_2 \text{ или } x_3; \\ L_3(\text{Объект4}) &= x_3.\end{aligned}$$

Для вычисления значения функции принадлежности L_3 (Объект2) воспользуемся операцией граничного объединения двух нечетких множеств $\mathbf{A} \oplus \mathbf{B}$ [9]. Результат определяется как нечеткое множество \mathbf{D} , заданное на том же универсуме \mathbf{U} , функция соответствия которого определяется по формуле

$$\mu_D(x) = \min\{\mu_A(x) + \mu_B(x), 1\}.$$

Тогда $\mu(x_1) \oplus (x_2) = \{0,02, 0,76, 1,0, 1,0, 1,0, 1,0, 0,51, 0,26, 0,02\}$.

Находим пересечения μ_{ans} с $\mu(x_1)$, $\mu(x_1) \oplus (x_2)$, $\mu(x_2)$, $\mu(x_3)$.

$$\begin{aligned}\mathbf{d}_1 &= \mu_{ans} \cap \mu(x_1) = \\ &= \{0,01, 0,20, 0,71, 0,75, 0,50, 0,26, 0,01, 0,01, 0,01\}; \\ \mathbf{d}_2 &= \mu_{ans} \cap (\mu(x_1) \oplus (x_2)) = \\ &= \{0,01, 0,20, 0,71, 1,0, 0,71, 0,3, 0,04, 0,005, 0,005\}; \\ \mathbf{d}_3 &= \mu_{ans} \cap \mu(x_2) = \\ &= \{0,01, 0,2, 0,50, 0,750, 0,71, 0,3, 0,04, 0,005, 0,005\}; \\ \mathbf{d}_4 &= \mu_{ans} \cap \mu(x_3) = \\ &= \{0,01, 0,01, 0,01, 0,255, 0,5, 0,3, 0,04, 0,005, 0,005\}.\end{aligned}$$

Максимальные значения составят:

$$\begin{aligned}\max(\mathbf{d}_1) &= 0,75; \\ \max(\mathbf{d}_2) &= 1,0; \\ \max(\mathbf{d}_3) &= 0,75; \\ \max(\mathbf{d}_4) &= 0,5.\end{aligned}$$

Выполняем коррекцию результирующих оценок по формуле (1)

$$Res_1 = 0,75; Res_2 = 0,32; Res_3 = 0,505; Res_4 = 0,42.$$

5. Пользователю предъявляется признак "Длина раковины"

$L_4 = (" \text{Длина раковины(ДР)}", \{x_1 = " \text{ДР у Объекта 1}", x_2 = " \text{ДР у Объекта 2}", x_3 = " \text{ДР у Объекта 3}", x_4 = " \text{ДР у Объекта 4} \})$,

$$U = [5,25], M = \{ \mu(x_1), \mu(x_2), \mu(x_3), \mu(x_4) \}.$$

Здесь функции принадлежности $\mu(x_j)$ имеют кусочно-линейную трапециевидную форму и задаются четверками $\langle a, b, c, d \rangle$ — абсциссами точек излома:

$$\mu(x_1) = \{8, 10, 12, 14\};$$

$$\mu(x_2) = \{12, 14, 16, 18\};$$

$$\mu(x_3) = \{12, 14, 15, 17\};$$

$$\mu(x_4) = \{10, 12, 15, 17\}.$$

Пользователь отвечает, что "Длина раковины равна примерно 11 мм".

Так как функции принадлежности представляют собой нечеткие интервалы, то ответ пользователя следует представить нечетким интервалом с треугольной функцией принадлежности (см. рисунок).

Ширина основания треугольника $a_{\max} - a_{\min}$ определяется степенью уверенности пользователя в ответе и зависит от его оценок: *уверен, не уверен, примерно, около, абсолютно уверен* и т. п.

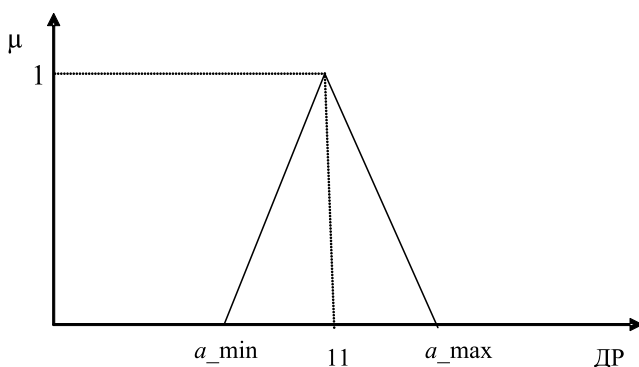
Значения a_{\max} и a_{\min} можно вычислить по формулам

$$a_{\max} = ans * (1 + w);$$

$$a_{\min} = ans * (1 - w),$$

где ans — числовое значение ответа пользователя (в примере оно равно 11), w — коэффициент не уверенности/уверенности пользователя.

Вычисление поправочного коэффициента d_i для каждого объекта сводится к нахождению пересечения двух нечетких интервалов: трапециевидного из базы данных признаков и треугольного, построенного по ответу пользователя.



Ответ пользователя задан нечетким интервалом

Для данного примера при $w = 0,3$:

$$d_1 = 1; d_2 = 0,43; d_3 = 0,43; d_4 = 0,81.$$

Тогда оценки с учетом ответа на вопрос о значении признака L_4 составят:

$$Res_1 = 0,75; Res_2 = 0,32; Res_3 = 0,43; Res_4 = 0,42.$$

Заключение

Предложенный подход представления знаний и организации нечеткого вывода в адаптивных таксономических экспертных системах реализован в серии компьютерных определителей. Авторами разработан прототип адаптивной экспертной системы "Таксакейс" (www.taxakeys.org). На ее основе совместно с группой экспертов созданы интерактивные определители представителей кладочер (Cladocera), копепоид (Copepoda), отрядов кормового зоопланктона и рыб семейства собачковых (Blenniidae) и сельдевых (Clupeidae) Севастопольского региона.

Исследование выполнено при финансовой поддержке РФФИ и Правительства города Севастополя в рамках научного проекта № 14-44-01566 "p_yug_a".

Авторы выражают благодарность В. В. Мельникову, С. А. Царину, А. В. Темных и М. И. Силакову за предоставленные материалы по конкретным группам объектов и помощь в наполнении базы знаний.

Список литературы

1. Лобанов А. Л., Кирейчук А. Г., Степаньянц С. Д., Смирнов И. С. Диагностические ключи: от текстовых дихотомических до компьютерных // Труды Зоологического института РАН. 2013. Приложение № 2. С. 249—268.
2. Adeeyo A. O., Ogunkunle A., Bello O. S. Review of Identification Aids in Biology: Past, Present and Future Challenges // Academic Journal of Interdisciplinary Studies MCSER Publishing, Rome-Italy. 2013. Vol. 2, N. 13. P. 27—34.
3. Dallwitz M. J. A Comparison of Interactive Identification Programs. URL: <http://delta-intkey.com/www/comparison/htm>
4. Lelekov S., Lyakh A. Taxex: Taxonomical Expert System — history of development and technology of identification / V. Berghe (Eds.) // Proceedings of "Ocean Biodiversity Informatics": an int. conf. on marine biodiversity data management. Hamburg, Germany, 29 Nov. — 1 Dec, 2004. Paris, UNESCO/IOC, VLIZ, BSH. 2007. P. 111—120.
5. Лях А. М., Лелеков С. Г. Адаптивная экспертная система для таксономической идентификации организмов // Экологическая безопасность прибрежной и шельфовой зон и комплексное использование ресурсов шельфа. Сб. научн. тр., Севастополь, 2014. Вып. 29. С. 107—113.
6. Лелеков С. Г., Лях А. М. Нечеткое описание признаков объектов в адаптивных таксономических экспертных системах // Вісник СевНТУ: зб. наук. пр. Серія: Автоматизація процесів та управління. Севастополь, 2014. Вип. 147/2014. С. 143—149.
7. Заде Л. А. Понятие лингвистической переменной и его применение к принятию приближенных решений. М.: Мир, 1976. 168 с.
8. Лелеков С. Г. Метод учета влияния модификаторов на нечеткие значения лингвистических переменных // Искусственный интеллект. 2013. № 4 (62). С. 454—462.
9. Леоненков А. В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. СПб.: БХВ, 2003. 736 с.

S. G. Lelekov, Associate Professor, e-mail: p1859@yandex.ru,
Sevastopol State University, Sevastopol,
A. M. Lyakh, Senior Scientist, e-mail: antonlyakh@yahoo.com,
Institute of biology of the Southern seas, Sevastopol

The Adaptive Fuzzy Inference System for Biological Taxonomy

Taxonomic identification of natural objects requires qualified researchers (experts) or qualitative identifiers which present expert knowledge. Realization of taxonomic identifiers in the form of computer expert systems leads to the needs to transform taxonomic diagnosis to formalized computer description. Diagnosis contain specific terminology and fuzzed statements that hamper their understanding and obstruct mathematical formalization. For that reason taxonomic expert systems not widely used.

Adaptive expert systems partially solved that problem because they takes into account semantic ambiguity of natural language. Adaptability means that system adapts to interpretation of indistinct user and expert concepts. The theory of fuzzy sets is the mathematical basis of adaptive expert systems functioning. According to algorithm presented in the paper the expert system transforms expert description of organism features into linguistic variables, presents them to the user for organism identification, transforms user answers into another linguistic variables, and compare user linguistic variable with expert ones to find a correspondence of organisms description stored in the system knowledge base with the user object.

The proposed approach of knowledge representation and organization of fuzzy inference in adaptive taxonomic expert systems is implemented in a series of computer identifiers. The authors have developed a prototype of an adaptive expert system "Taxakeys" (taxakeys.org). On its basis, together with the group of experts, the interactive identifiers of Cladocera, Copepoda, zooplankton and two fish families Blennidae and Clupeidae of the Sevastopol region have been created.

Keywords: expert systems, linguistic variables, fuzzy inference, biological taxonomy, computer taxonomic keys

References

1. **Lobanov A. L., Kirejchuk A. G., Stepan'janc S. D., Smirnov I. S.** Diagnosticheskie kljuchi: ot tekstovyh dihotomicheskikh do komp'yuternyh, *Trudy Zoologicheskogo instituta RAN*, 2013, Prilozhenie, no. 2, pp. 249—268.
2. **Adeeyo A. O., Ogunkunle A., Bello O. S.** Review of Identification Aids in Biology: Past, Present and Future Challenges. *Academic Journal of Interdisciplinary Studies MCSEER Publishing*, Rome-Italy, 2013, vol. 2, no. 13, pp. 27—34.
3. **Dallwitz M. J.** *A Comparison of Interactive Identification Programs*. URL: <http://delta-intkey.com/www/comparison/htm>
4. **Lelekov S., Lyakh A.** Taxex: Taxonomical expert system — history of development and technology of identification. V. Berghe (eds.). *Proceedings of "Ocean Biodiversity Informatics": an Int. conf. on marine biodiversity data management*. Hamburg, Germany, 29 Nov. — 1 Dec, 2004. Paris: UNESCO/IOC, VLIZ, BSH, 2007, pp. 111—120.
5. **Lyakh A. M., Lelekov S. G.** Adaptivnaja ekspertnaja sistema dlja taksonomicheskoi identifikacii organizmov. *Ekologicheskaja bezopasnost pribrezhnoj i shelfovoj zon i kompleksnoe ispolzovanie resursov shelfa*. Sb. nauchn. tr. Sevastopol, 2014, iss. 29, pp. 107—113.
6. **Lelekov S. G., Lyakh A. M.** Nechetkoe opisanie priznakov ob'ektov v adaptivnyh taksonomicheskikh jekspertnyh sistemah. *Visionik SevNTU: zb. nauk. pr. Serija: Avtomatizacija procesiv ta upravlinja*. Sevastopol, 2014, iss. 147/2014, pp. 143—149.
7. **Zade L. A.** *Ponjatje lingvisticheskoj peremenoj i ego primenenie k prinjatiju priblizhennyh reshenij*. Moscow: Mir, 1976. 168 p.
8. **Lelekov S. G.** Metod ucheta vlijanija modifikatorov na nechetkie znacheniiia lingvisticheskikh peremennyh, *Artificial intelligence*, 2013, no. 4 (62), pp. 454—462.
9. **Leonenkov A. V.** Nechetkoe modelirovanie v srede MATLAB i fuzzy TECH. SPb.: BHV, 2003, 736 p.

УДК 004.272.34

В. С. Дыкин, магистрант, **В. Ю. Мусатов**, канд. техн. наук, доцент, **А. С. Варезников**, аспирант,
А. А. Большаков, д-р техн. наук, профессор, **В. В. Сысоев**, д-р техн. наук, профессор
СГТУ, г. Саратов, e-mail: robsii@sstu.ru

Применение генетического алгоритма для определения архитектуры нейронной сети для прибора вида "электронный нос"

Рассматривается возможность применения генетических алгоритмов для выбора топологии искусственных нейронных сетей в системе распознавания газоаналитического устройства вида "электронный нос". Поиск топологии осуществлен оптимизацией по критериям увеличения процента правильного распознавания и уменьшения вычислительных затрат с использованием массива заранее заданных вероятных значений параметров архитектуры нейронной сети. Оптимизируемые характеристики: число скрытых слоев, количество нейронов в каждом слое и объем обучающей выборки. Распознавание проведено в системе Matlab нейронной сетью прямого распространения с функцией обучения Левенберга-Марквардта. Полученные результаты подтверждают работоспособность созданного алгоритма поиска топологии искусственной нейронной сети.

Ключевые слова: "электронный нос", генетические алгоритмы, мультисенсорная микросистема, анализ газов, распознавание образов