РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ МАТЕМАТИКИ имени С,Л,СОБОЛЕВА

МЕТОДЫ АНАЛИЗА ДАННЫХ

(Вычислительные системы, 111)
Сборник научных трудов

НОВОСИБИРСК 1985

Стр. 23-32

==

КОНСТРУКТИВНОЕ ЧИСЛОВОЕ ПРЕДСТАВЛЕНИЕ ВЕЛИЧИН

Витяев Е.Е.

Введение.

Исследования, проводимые в психологии, социологии, принятии решений, экспертном оценивании и других областях, показывают, что есть много сложных, структурных «нечисловых» величин (частичные порядки, толерантности, решетки и т.д.). Логический анализ таких величин, проведенный в теории измерений [1,2], теории принятия решений [3,4] и анализе нечисловой информации [5-7], показал, что формальные представления таких величин – эмпирические системы – являются такими алгебраическими структурами, которые нельзя сильным гомоморфизмом отобразить в поле вещественных чисел, т.е. для таких величин нельзя построить их числовые представления в теории измерений. С другой стороны, числовые представления величин обладают следующими достоинствами: они «удобны», по числовым значениям величин легко определяются исходные (в эмпирической системе) соотношения между значениями величин, для числовых величин разработано много математических методов их обработки. Поэтому, наряду с необходимостью разрабатывать «прямые» (например, логические) методы обработки структурных «нечисловых» величин, остается важной задача построения их числовых представлений. Именно этим мотивировалось включение этой задачи в программную систему логического анализа данных ЛАДА (см. наст. сб.).

Рассмотрим смысл и роль числового представления. Смысл состоит в том, чтобы значениям величины приписать числа так, чтобы исходные отношения и операции преобразовывались в некоторые «простые» и «удобные» числовые отношения и операции. В этом случае по значениям числовых отношений и операций легко определяются значения исходных отношений и операций.

В теории измерений числовое представление строится по следующей схеме. Предполагается, что величины измеряются некоторыми конкретными приборами. Свойства величин исследуются экспериментально и фиксируются в законах и гипотезах. Эмпирические системы величин являются идеализациями

соответствующих приборов и процедур их использования. Величины известны нам в той мере, в какой известны соответствующие законы и гипотезы. Формально законы и гипотезы задаются системами аксиом, поэтому относительно эмпирической системы известно только, что они удовлетворяют соответствующей системе аксиом. Числовые представления величин получаются подбором такой числовой системы, в которую сильным гомоморфизмом отображается любая эмпирическая система, удовлетворяющая системе аксиом, описывающей данную величину.

Построенные по такой схеме числовые представления обладают следующими недостатками. В качестве числовых отношений и операций используется небольшое число математических действий. Этого достаточно для числового представления большинства числовых величин [2], но это препятствует числовому представлению многих других величин. Доказательство, что любая эмпирическая система, удовлетворяющая системе аксиом, сильным гомоморфизмом отображается в выбранную числовую систему, предъявляет чрезмерно сильные требования к системе аксиом. Приходиться включать в нее аксиомы, не поддающиеся экспериментальной проверке, а также «чисто технические» аксиомы, не изменяющие множества экспериментально проверяемых следствий [1]. Это противоречит содержанию систем аксиом, как результатам экспериментального анализа свойств величин. Такие аксиомы часто отражают свойства числовой системы, а не свойства величин. Чрезмерным является также требование сильного гомоморфизма любой (?) алгебраической системы в числовую систему. Если теорию T (систему аксиом) рассматривать как типы данных, а эмпирическую систему как структуру данных, то числовое представление строиться для любой структуры данного типа T. В программировании [8] тем не менее рассматриваются и другие способы задания семантики типа данных – инициальные, финальные, простые и другие структуры. Поскольку эти структуры определенным образом связаны со всеми остальными структурами типа T, то числовое представление можно, в принципе, строить только для этих структур. Таким путем может быть получено конструктивное числовое представление экстенсивных величин. С практической точки зрения также имеет смысл получать числовые представления не для всех структур, а только для конечных, конечно-порожденных или конечных подструктур структур типа T. Отмеченные ограничения в постановке проблемы существования числового представления влекут за собой и соответствующие ограничения в постановках других проблем (проблем единственности и адекватности).

Смыслу числового представления точнее всего соответствует понятие конструктивизации [9] эмпирической системы. В этом случае значениям величины приписываются натуральные, рациональные или другие числа (или коды) так, чтобы значения отношений и операций в эмпирической системе можно было эффективно определить по этим числам. Такой способ получения числовых представлений не накладывает на числовые отношения и операции никаких ограничений, кроме эффективности, предъявляет более слабые требования к системе аксиом и не связан с требованием существования гомоморфизма в какие-то другие системы. Этот способ называется в работе конструктивным числовым представлением и может использоваться для числового представления структурных «нечисловых» величин. Заметим, что для конструктивных числовых представлений справедливы некоторые из замечаний, сделанных относительно постановки основных проблем (существования, единственности и адекватности)

в теории измерений. Их учет открывает широкие возможности для дальнейших исследований.

1. **Основные понятия теории измерений**.

Пусть знания о некоторой величине, свойстве, признаке сформулированы в некоторой теории T сигнатуры $\Omega = \langle P_0,P_1,\ldots,P_n, \rho_1,\ldots,\rho_m, c_0,c_1,c_2,\ldots\rangle$, где P_i , $i\leq n$ – предикатные символы; ρ_j , $j\leq m$ – символы операций; c_l , $l \in I$ – символы констант ($I = \varnothing$, I – начальная часть ряда натуральных чисел $\omega = \{0,1,2,\ldots\}$, $I = \omega$); P_0 – равенство. Величиной будем называть неприводимую [1] (равенство является единственным отношением конгруентности) систему $\Im = \langle A; \Omega_\Im\rangle$ сигнатуры Ω, удовлетворяющую теории T; A – множество значений величины, $\Omega_\Im = \{P^\Im_0,P^\Im_1,\ldots,P^\Im_n, \rho^\Im_1,\ldots,\rho^\Im_m, c^\Im_0, c^\Im_1, c^\Im_2,\ldots\}$ – множество отношений, операций и констант типа Ω, интерпретируемых в понятиях предметной области. Числовыми системами называются системы $\Re = \langle Re^k, \Omega_R\rangle$ сигнатуры Ω, где к – размерность числового представления, Re – поле вещественных чисел, $\Omega_\Re = \{=,P^\Re_1,\ldots,P^\Re_n, \rho^\Re_1,\ldots,\rho^\Re_m, c^\Re_0, c^\Re_1, c^\Re_2,\ldots\}$ – множество отношений, операций и констант, определенных на Re или Re^k. Зафиксируем некоторую систему $\Re$. Шкалой [1] (числовым представлением) величины $\Im = \langle A; \Omega_\Im\rangle$ называется отображение (сильный гоморфизм) μ: A $\to Re^k$, удовлетворяющее условиям:

1. $P^\Im_i(a_1,\ldots,a_{mi}) \Leftrightarrow P^\Re_i(\mu a_1,\ldots, \mu a_{mi})$, $i = 0,1,\ldots,n$;
2. $\rho^\Im_j (a_1,\ldots,a_{mj}) \Leftrightarrow \rho^\Re_j(\mu a_1,\ldots, \mu a_{mj})$, $j = 1,\ldots,m$;
3. $\mu c^\Im_l = c^\Re_l$, $l \in I$.

Сильный гомоморфизм μ: $\Im \to \Re$ изоморфно отображает величину $\Im$ в числовую систему $\Re$. Введем обозначения. AC(T) – множество неприводимых (алгебраических) систем теории T; $AC_\aleph(T)$, $AC_\omega(T)$ – подмножества AC(T), содержащие системы не более чем континуальной и счетной мощности соответственно; $F(\Im,\Re)$ – множество шкал величины $\Im$.

В теории измерений исследуются три основные проблемы [1,2].

1. **Проблема существования**. Для данной теории T величины найти достаточно простую и удобную числовую систему $\Re$ (например, поле вещественных чисел) и доказать, что для любой величины $\Im \in AC_\aleph(T)$ существует шкала $(F(\Im,\Re) \neq \varnothing\}$. Из формулировки проблемы существования следует, что знаний T должно быть достаточно для выбора числовой системы $\Re$ и построения шкалы для любой системы $\Im \in AC_\aleph(T)$. Системы из $AC_\aleph(T)$ являются величинами, которые удовлетворяют нашим знаниям T о них и для которых мы можем построить числовое представление. Решение проблемы существования должно, кроме того, давать метод шкалирования приборов, измеряющих эти величины. Этот метод обычно извлекается из доказательства теоремы существования.

2. **Проблема единственности**: Для выбранной числовой системы $\Re$ определить все шкалы $F(\Im,\Re)$ величин $\Im \in AC_\aleph(T)$. Эти множества можно, в частности, определить, найдя группу допустимых преобразований [1].

Обычно требуется, чтобы не только числовая система, но и все множества $F(\Im,\Re)$ были просты и удобны. Простота и удобство нужны для решения следующей проблемы.

3. **Проблема адекватности**. Числовые утверждения должны быть инвариантны относительно произвола в выборе шкал из $F(\Im,\Re)$ (см. [1]).

Решение этих проблем позволяет корректно вводить числовые представления величин и в определенной степени корректно их использовать.

2. **Конструктивные числовые представления величин**.

При конструктивном представлении величин значения $a \in A$ величин $\Im = \langle A; \Omega_{\Im} \rangle \in AC_{\omega}(T)$ нумеруются (кодируются). Нумерацией множества A называется отображение ν множества натуральных чисел $\omega = \{0,1,2,\ldots\}$ на A, $\nu: \omega \to A$ [9]. Пару $(\Im,\nu)$ будем называть конструктивным представлением величины $\Im$ (конструктивной системой [9]), а нумерацию ν – конструктивным числовым представлением (конструктивизацией [9]), если существует характеристические общерекурсивные функции $P^N_0,P^N_1,\ldots,P^N_n$ со значениями во множестве $\{0,1\}$, общерекурсивные функции $\rho^N_1,\ldots,\rho^N_m$ и натуральные числа c^N_0, c^N_1, $c^N_2,\ldots$ такие, что

1. $P^{\Im}_i(\nu a_1,\ldots, \nu a_{mi}) \Leftrightarrow (P^N_i(a_1,\ldots,a_{mi}) = 1)$, $i = 0,1,\ldots,n$;
2. $\rho^{\Im}_j(\nu a_1,\ldots, \nu a_{mj}) \Leftrightarrow \nu\rho^N_j (a_1,\ldots,a_{mj})$, $j = 1,\ldots,m$;
3. $c^{\Im}_l = \nu c^N_l$, $l \in I$.

Конструктивное числовое представление ν аналогично шкале, только вместо числовых отношений, операций и констант используются общерекурсивные функции и натуральные числа. Конструктивной числовой системой является система $N = \langle\omega;\Omega_N\rangle$, $\Omega_N = \{P^N_0,P^N_1,\ldots,P^N_n, \rho^N_1,\ldots,\rho^N_m, c^N_0, c^N_1, c^N_2,\ldots\}$ Сформулируем проблемы существования, единственности и адекватности для конструктивного числового представления.

2.1. **Проблема существования. 1**. Доказать, что для любой системы N существует конструктивное числовое представление.

Данная формулировка предъявляет довольно сильные требования к теории T. Более слабой, но также практически интересной является следующая формулировка проблемы существования. Пусть Ф – система аксиом теории T, $Ф^*$ – совокупность эрбрановых форм предложений Ф (скулемизация Ф [9, с.345]), $f_1,f_2,\ldots$ – символы скулемовских функций. Определим сигнатуру $\Omega^* = \Omega\cup\{f_1,f_2,\ldots\}$. Скулемовской оболочкой $\Im^*(X)$ подмножества $X \subset |\Im^*|$ системы $\Im^* \in AC(Ф^*)$ сигнатуры Ω^* называется подсистема системы $\Im^*$, порожденная подмножеством X. Можно доказать [9], что $\Im^*(X) \in AC(Ф^*)$ для любого подмножества $X \subset |\Im^*|$.

Проблема существования. 2. Доказать, что для любой величины $\Im^* \in AC(Ф^*)$ сигнатуры Ω^* и любого **конечного** подмножества $X \subset |\Im^*|$ скулемовская оболочка $\Im^*(X)$ имеет конструктивное числовое представление.

2.2. **Проблема единственности**. Ее можно разбить на две части. Первая связана с существованием не сводимых друг к другу посредством эффективного отображения (неавтоэквивалентных [9]) конструктивных числовых представлений. В [10] показано, что число неавтоэквивалентных конструктивизаций может быть любым. Неавтоэквивалентные конструктивные числовые представления принципиально различны, поэтому необходимо учитывать возможный произвол в выборе одной из них.

Проблема единственности A. Для каждой величины $\Im = \langle A; \Omega_{\Im}\rangle \in AC_{\omega}(T)$ определить число неавтоэквивалентных конструктивных числовых представлений.

Вторая часть проблемы единственности так же, как и в случае числовых представлений, связана с произволом в выборе одного из автоэквивалентных конструктивных числовых представлений.

Проблема единственности Б. Для каждой величины $\Im$ = <A; $\Omega_{\Im}$> $\in$ $AC_{\omega}(T)$ определить все классы автоэквивалентных конструктивных числовых представлений.

2.3. **Проблема адекватности**. Она также разбивается на две части в зависимости от того, какой произвол в выборе конструктивных числовых представлений нужно учитывать.

Проблема адекватности А. Выбор класса автоэквивалентных числовых представлений должен учитывать имеющиеся знания Т.

Проблема адекватности Б. Числовые утверждения должны быть инвариантны относительно выбора одного из автоэквивалентных конструктивных числовых представлений.

3. Взаимосвязь конструктивного и числового представлений.

Предположим, что для некоторой величины, описываемой теорией Т, решены проблемы существования, как для числового, так и конструктивного числового представлений. Пусть $\Re$ – выбранная числовая система, $\Im \in AC_{\aleph}(T)$ – величина и μ – шкала. Из решения проблемы существования конструктивного числового представления следует, что для любой не более чем счетной величины $\Im_{\omega} \in AC_{\omega}(T)$ являющейся подсистемой величины $\Im$ ($\Im_{\omega} \subset \Im$) существует конструктивное числовое представление ν: $\omega \to |\Im_{\omega}|$. Рассмотрим образ $\mu(\Im_{\omega})$ величины $\Im_{\omega}$ при его отображении в числовую систему $\Re$. Так как подсистема $\Im_{\omega}$ содержит все константы $c^{\Im}_{l}$, $l \in I$ и замкнута относительно операций, то из определения шкалы следует, что отображение μ: $\Im_{\omega} \to \Re$ также является шкалой величины $\Im_{\omega}$. Поэтому для каждой подсистемы $\Im_{\omega} \in AC_{\omega}(T)$ любой из величин $\Im \in AC_{\aleph}(T)$, $\Im_{\omega} \subset \Im$ существует, как конструктивное ν, так и числовое μ представление величины. Рассмотрим отображение $\mu\nu$: $\omega \to \mu(\Im_{\omega})$. Из определений шкалы и конструктивного числового представления следует, что пара $(\mu(\Im_{\omega}),\mu\nu)$ является конструктивным представлением числового представления $\mu(\Im_{\omega})$ величины ($\Im_{\omega}$). Поэтому для величин $\Im_{\omega} \in AC_{\omega}(T)$, $\Im_{\omega} \subset \Im$, $\Im \in AC_{\aleph}(T)$ существуют конструктивное числовое представление ν числовое представление μ и конструктивное представление $\mu\nu$ числового представления $\mu(\Im_{\omega})$.

4. Примеры конструктивных представлений величин.

3.1. Рассмотрим линейный порядок. Знания Т о линейном порядке содержат аксиомы антисимметричности, полноты и транзитивности. Линейными порядками являются, например, балльные величины и величины типа «число»: число рабочих на предприятии, число автокатастроф, число браков или разводов и т.д. Значениями многих таких величин являются натуральные числа, поэтому их естественным числовым представлением может быть конструктивное числовое представление. Такие величины удовлетворяют следующей аксиоме.

Аксиома 1. Любая ограниченная сверху (снизу) последовательность $a_1 < a_2 < a_3 < \ldots < a$ ($a < \ldots < a_3 < a_2 < a_1$) конечна.

Обозначим теорию линейного порядка Т вместе с аксиомой 1 через T_1. Известно, что любой линейный, но не более чем счетный порядок, удовлетво-

ряющий теории T_1, вложим в модель $<\omega; \leq>$. Отсюда следует решение проблемы существования конструктивного числового представления для линейных порядков T_1.

Предложение 1. Любой линейный порядок $\Im \in AC_\omega(T_1)$ имеет конструктивное числовое представление.

Конструктивными числовыми представлениями могут служить обычные способы нумерации значений этих величин.

Рассмотрим линейные порядки, удовлетворяющие аксиоме полноты.

Аксиома 2. $\forall a,b,\exists c(a < c < b)$.

Обозначим через T_2 теорию линейного порядка T вместе с аксиомой 2. Примерами полных линейных порядков, удовлетворяющих T_2, являются физические величины, используемые в нефизических областях. Например, величины температуры, давления, веса человека, рассматриваемые с медицинской точки зрения, или температуры, освещенности, влажности почвы, рассматриваемые с сельскохозяйственной точки зрения. Для этих величин операция сложения (имеющая смысл с физической точки зрения) смысла не имеет. Осмысленно только отношение порядка, являющееся полным линейным порядком. Такой порядок естественно представлять не действительными, а рациональными числами. Получим конструктивное числовое представление полных линейных порядков, используя рациональные числа. Известно, что любой полный, не более чем счетный линейный порядок $\Im \in AC_\omega(T_2)$ изоморфен одному из интервалов (0,1), [0,1), (0,1], [0,1] множества рациональных чисел.

Предложение 2. Любой полный линейный порядок $\Im \in AC_\omega(T_2)$ имеет конструктивное числовое представление.

Примерами конструктивных числовых представлений могут служить градации шкал приборов, измеряющих эти величины.

Рассмотрим деревья – рефлексивные, антисимметричные, транзитивные порядки, удовлетворяющие следующей аксиоме.

Аксиома 3. $\forall a,b,c(c \leq a \,\&\, c \leq b \Rightarrow a \leq b \lor b \leq a)$

Обозначим теорию деревьев через T_3. Конечными деревьями описываются такие величины, как должность, занимаемое место (в дереве рабочих мест некоторой организации), иерархические величины и т.д. Конечные деревья всегда конструктивизируемы, поэтому решение проблемы существования конструктивного числового представления сводиться к построению простой и удобной конструктивизации, применимой к любому конечному дереву. Пример такого конструктивного числового представления приведен на рисунке 1.

Если у дерева несколько корневых вершин, то они нумеруются числами 1,2,3,... Вершинам дерева (значениям величины) сопоставляются наборы натуральных чисел $a = \nu(<n^a{}_1,\ldots, n^a{}_k>)$, $b = \nu(<n^b{}_1,\ldots, n^b{}_m>)$. По числам из набора легко определяется отношение порядка между a и b.

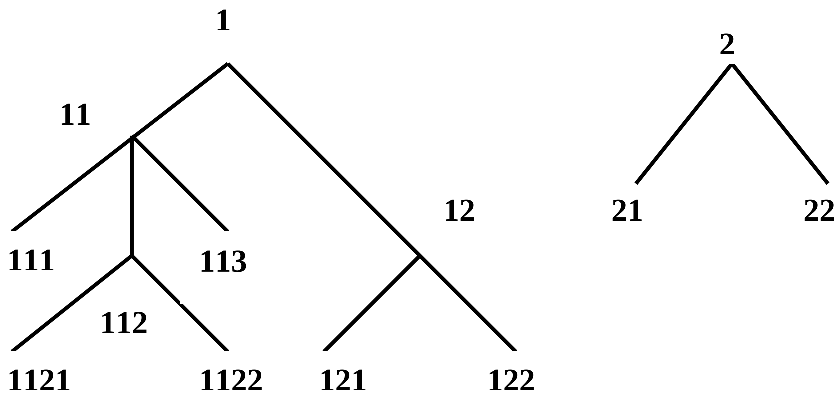

рис 1.

Закономерные связи между конструктивными числовыми представлениями также должны задаваться общерекурсивными функциями. Например, закономерная связь между древовидными величинами – занимаемое место, должность; балльными величинами – степень, звание; полно линейно упорядоченной величиной – зарплата, определяется законодательством и может быть представлена общерекурсивной функцией.

5. Конструктивные измерительные процедуры, тесты и анкеты.

Шкалы μ: $\Im \to \Re$ практически реализуются в виде шкал приборов (весов, линейки, термометра). Конструктивизации ν также могут реализовываться как показания некоторых измерительных процедур, в частности тестирования, анкетирования, обследования и т.д. Предположим, что нас интересует отношение предпочтения некоторой величины $\Im = \langle A; \leq \rangle$ (коэффициента интеллектуальности, удовлетворенности работой, температура) и способ прямого измерения отношения предпочтения $\leq$ дорог, неудобен, требует большого времени и т.д. Для более простого и быстрого измерения этого отношения разрабатывается и используется тест (анкета, обследование). Применение теста к испытуемому (респонденту, больному) дает для некоторого значения $a \in A$ величины $\Im$ набор ответов в виде некоторой последовательности натуральных или рациональных чисел $\langle n^a_1, \ldots, n^a_k \rangle$. Если по результатам теста для любых двух значений $a, b \in A$ величины $\Im$ можно эффективно определить отношение предпочтения

$$a \leq b \Leftrightarrow P^N(\langle n^a_1, \ldots, n^a_k \rangle, \langle n^b_1, \ldots, n^b_l \rangle)$$

например, подсчитывая сумму баллов, взвешенное среднее, кодируя ответы и т.д.), то отображение $\Im$: $\langle n^a_1, \ldots, n^a_k \rangle \to a$, осуществляемое тестом, и будет конструктивным числовым представлением величины $\Im = \langle A; \leq \rangle$. Сама процедура тестирования (анкетирования, обследования) будет конструктивной измерительной процедурой со значениями вида $\langle n^a_1, \ldots, n^a_k \rangle$.

Примером такого отношения предпочтения и соответствующего теста является отношение предпочтения между односемейными домами [11,с.243].

Использование теста для конструктивного измерений некоторой величины может быть обосновано решением следующей задачи. Пусть задано некоторое отношение предпочтения $\leq$ величины $\Im = \langle A; \leq \rangle$, обладающее свойствами T (удовлетворяющее аксиомам частичного порядка, толерантности, решеток). Требуется решить проблему существования конструктивного числового представления ν для любой величины $\Im \in AC_\omega(T)$ и затем для данной нам величины

$\mathfrak{I}$ разработать тест, измеряющий ее конструктивное числовое представление ν. Мы не можем сразу строить конструктивное числовое представление ν для исходной величины $\mathfrak{I}$, так как она известна нам только своими аксиомами, содержащимися в Т. Поэтому решить проблему существования конструктивного числового представления нужно опираясь на $AC_{\omega}(T)$.

В работе получены конструктивные числовые представления лишь для простейших величин. Конструктивные числовые представления конечных решеток получены В.Ф.Новиковым. Дальнейшее построение конструктивных числовых представлений может быть стимулировано применением программной системы логического анализа данных ЛАДА.

ЛИТЕРАТУРА

1. Пфанцагль И. Теория измерений. М., Мир, 1976, с.248.
2. Krantz D.H., Luce R.D., Suppes P., Tversky A. Foundations of measurement. Vol. 1,2,3. - NY, London: Acad. press, 1971, 1989, 1990.
3. Фишберн П. Теория полезности для принятия решений. М, Наука, 1978, с.352.
4. Кини Р.Л., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения. М. Радио и связь, 1981, с.560.
5. Анализ нечисловой информации // ТюринЮ.Н., Литвак Б.Г., Оргов А.И. и др. М., 1981. с.80 (Препринт // Научный совет по комплексной проблеме «Кибернетика»).
6. II Всесоюзная школа-семинар «Программно-алгоритмическое обеспечение прикладного многомерного статистического анализа». (Тезисы докладов). М., 1983. с.374
7. Всесоюзная конференция «Нечисловая статистика, экспертные оценки и смежные вопросы» (Тезисы докладов). М.-Таллин, 1984, 403с.
8. Данные в языках программирования. Абстракция и типология. М., Мир, 1982, с.327.
9. Ершов Ю.Л. Проблемы разрешимости и конструктивные модели. М., Наука, 1980, с.415.
10. Гончаров С.С. Проблема числа неавтоэквивалентных конструктивизаций. Алгебра и Логика, 19, №6, 1980, с.621-639.
11. Козелецкий Ю. Психологическая теория решений. М., Прогресс, 1979, с.504.