УДК 303.732.4

МАТЕМАТИЧЕСКАЯ СУЩНОСТЬ СИСТЕМНОЙ ТЕОРИИ ИНФОРМАЦИИ (СТИ) (Системное обобщение формулы Больцмана-Найквиста-Хартли, синтез семантической теории информации Харкевича и теории информации Шеннона)

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор

Кубанский государственный аграрный университет, Краснодар, Россия

В статье кратко описана математическая сущность предложенной автором системной теории информации (СТИ), являющейся математической моделью системно-когнитивного анализа (СК-анализ) реализуемой в его программном инструментарии – универсальной когнитивной аналитической системе "Эйдос"

Ключевые слова: СИСТЕМНАЯ ТЕОРИЯ ИНФОРМАЦИИ, МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ,СК-АНАЛИЗ, ПРОГРАММНЫЙ ИНСТРУМЕНТАРИЙ, УНИВЕРСАЛЬНАЯ КОГНИТИВНАЯ АНАЛИТИЧЕСКАЯ СИСТЕМА "ЭЙДОС".

UDC 303.732.4

MATHEMATICAL ESSENCE OF SYSTEMIC THEORY OF INFORMATION(STI) (Systemic substantiation of Boltsman-Naikequest-Hartly formula, synthesis of semantic information theory of Kharkevich and information theory of Shennon)

Lutsenko Evgeny Veniaminovich
Dr. Sci.Econ., Cand. Tech.Sci., professor

Kuban State Agrarian University, Krasnodar, Russia

The mathematical essence of systemic information theory(STI) offered by the author is the mathematic model of systemic-cognitive analysis (SC-analysis) realized in its programming instruments – universal cognitive analytical system “AIDOS”.

Key words: SYSTEMIC THEORY OF INFORMATION, MATHEMATIC-COGNITIVE ANALYSIS, SC-ANALYSIS, PROGRAMMING INSTRUMENTS, UNIVERSAL COGNITIVE NALYTICAL SYSTEM “ AIDOS”.

В работах [2, 4, 5, 12, 13, 14] предложена и развита *программная идея системного обобщения математики путем тотальной замены понятия множества на более общее и содержательное понятие системы и тщательное прослеживания всех последствий этого во всех понятиях математики, основанных на понятии множества*. Частично эта идея была реализована при разработке автоматизированного системно-когнитивного анализа (АСК-анализа) [2], математическая модель которого основана на системном обобщении формул для количества информации Найквиста-Хартли и Харкевича, в результате чего удалось *объединить* семантическую теорию информации Харкевича с теорией информации Шеннона. Полученная в результате этого обобщенная теория названа системной теорией информации (СТИ). Системное обобщение понятия функциональной зависимости рассматривается в работах [7 19], в них же вводятся новые научные понятия и соответствующие термины: "когнитивные функции" и "когнитивные числа". Различные аспекты СТИ рассмотрены в работах [1-21]. Таким образом существует много работ, в том числе монографий и учебных пособий, в которых изложена системная теория информации. Однако, как раз то, что этих работ много и они довольно объемны не вдохновляет потенциальных читателей.

Поэтому в данной статье ставится **цель** *описать математическую суть системной теории информации на столько кратко, на сколько это* *вообще возможно*. Своеобразной "ценой" за это является то, что пришлось пожертвовать подробной смысловой интерпретацией полученных математических выражений, за которой мы отсылаем к уже упомянутым работам.

Итак, классическая формула Хартли имеет вид:

$$I = Log_2 W \tag{1}$$

Будем искать ее системное обобщение в виде:

$$I = Log_2 W^j \tag{2}$$

где:

W – количество элементов в множестве.

j – коэффициент эмерджентности, названный автором в честь Хартли коэффициентом эмерджентности Хартли.

Примем, что системное обобщение формулы Хартли имеет вид:

$$\boxed{I = Log_2 \sum_{m=1}^{M} C_W^m} \tag{3}$$

где:

C_W^m – количество подсистем и m элементов;

m – сложность подсистем;

M – ***максимальная*** сложность подсистем.

Так как $C_W^1 = W$, то при $M=1$ система переходит в множество и выражение (3) приобретает вид (1), т.е. для него выполняется *принцип соответствия*, являющийся обязательным для более общей теории.

Учитывая, что при $M=W$:

$$\sum_{m=1}^{M} C_W^m = 2^W - 1 \tag{4}$$

в этом случае получаем:

$$I = Log_2 (2^W - 1) \tag{5}$$

Выражение (5) дает *приближенную оценку максимального количества информации* в элементе системы. Из выражения (5) видно, что при увеличении числа элементов W количество информации I быстро стремится к W (6) и уже при $W>4$ погрешность выражения (5) не превышает 1%:

$$\begin{gathered} при \;\; W \to \infty \\ I \to W \end{gathered} \tag{6}$$

Приравняв правые части выражений (2) и (3):

$$I = Log_2 W^{j} = Log_2 \sum_{m=1}^{M} C_W^m \qquad (7)$$

получим выражение для коэффициента эмерджентности Хартли:

$$\boxed{j = \frac{Log_2 \sum_{m=1}^{M} C_W^m}{Log_2 W}} \qquad (8)$$

Смысл этого коэффициента раскрыт в работах [2, 4, 5, 9, 12, 13, 14]. Здесь отметим лишь, что при *M*®*1* когда система асимптотически переходит в множество *j* ®*1* и (2) ® (1), как и должно быть согласно принципу соответствия.

С учетом (8) выражение (2) примет вид:

$$I(W,M) = Log_2 W^{\frac{Log_2 \sum_{m=1}^{M} C_W^m}{Log_2 W}} \qquad (9)$$

или при $M=W$ и больших W, учитывая (4 и 5):

$$I(W,M) = Log_2 W^{\frac{W}{Log_2 W}} = W \qquad (10)$$

Выражение (9) и представляет собой искомое системное обобщение классической формулы Хартли, а выражение (10) – его достаточно хорошее приближение при большом количестве элементов в системе W.

Классическая формула А.Харкевича имеет вид:

$$I_{ij}(W,M) = Log_2 \frac{P_{ij}}{P_j} \qquad (11)$$

где: – P_{ij} – условная вероятность перехода объекта в j-*е* состояние *при условии* действия на него i-*го* значения фактора;

– P_j – безусловная вероятность перехода объекта в j-*е* состояние (вероятность самопроизвольного перехода или вероятность перехода, посчитанная по всей выборке, т.е. при действии *любого* значения фактора).

Придадим выражению (11) следующий ***эквивалентный*** вид, который и будем использовать ниже:

$$I_{ij}(W,M) = Log_2 \frac{P_{ij}}{P_i} \qquad (12)$$

где: – индекс i обозначает признак (значение фактора): *1£ i £ M*;

– индекс j обозначает состояние объекта или класс: *1£ j £ W*;

– P_{ij} – условная вероятность наблюдения *i-го* значения фактора у объектов в *j-го* класса;

– P_i – безусловная вероятность наблюдения *i-го* значения фактора по всей выборке.

Из (12) видно, что *формула Харкевича для семантической меры информации по сути является логарифмом от формулы Байеса для апостериорной вероятности (отношение условной вероятности к безусловной).* Вопрос об эквивалентности выражений (11) и (12) рассмотрим позднее.

Известно, что классическая формула Шеннона для количества информации для неравновероятных событий преобразуется в формулу Хартли при условии, что события равновероятны, т.е. удовлетворяет фундаментальному *принципу соответствия*. Поэтому теория информации Шеннона справедливо считается обобщением теории Хартли для неравновероятных событий. Однако, выражения (11) и (12) при подстановке в них реальных численных значений вероятностей P_{ij}, P_j и P_i не дает количества информации в ***битах***, т.е. для этого выражения не выполняется *принцип соответствия*, обязательный для более общих теорий. Возможно, в этом состоит причина довольно сдержанного, а иногда и скептического отношения специалистов по теории информации Шеннона к семантической теории информации Харкевича.

Причину этого мы видим в том, что в выражениях (11) и (12) отсутствуют глобальные параметры ***конкретной*** модели W и M, *т.е. в том, что* А.Харкевич в своем выражении для количества информации не ввел зависимости *от мощности пространства будущих состояний объекта W и количества значений факторов M*, обуславливающих переход объекта в эти состояния.

Поставим задачу получить такое обобщение формулы Харкевича, которое бы удовлетворяло ***тому же самому*** *принципу соответствия, что и формула Шеннона, т.е. преобразовывалось в формулу Хартли в предельном детерминистском равновероятном случае, когда каждому классу (состоянию объекта) соответствует один признак (значение фактора), и каждому признаку – один класс, и эти классы (а, значит и признаки),* ***равновероятны,*** *и при этом каждый фактор однозначно, т.е.* ***детерминистским*** *образом определяет переход объекта в определенное состояние, соответствующее классу.*

Будем искать это обобщение (12) в виде:

$$I_{ij}(W,M) = Log_2\left(\frac{P_{ij}}{P_i}\right)^{\Psi} \qquad (13)$$

Найдем такое выражение для коэффициента Ψ, названого нами в честь А.Харкевича "коэффициентом эмерджентности Харкевича", которое обеспечивает выполнение для выражения (13) принципа соответствия с классической формулой Хартли (1) и ее системным обобщением (2 и 3) в *равновероятном детерминистском* случае.

Для этого нам потребуется выразить вероятности P_{ij}, P_j и P_i через частоты наблюдения признаков по классам (см. таблица 1). В таблице 1 рамкой обведена область значений, переменные определены ранее.

Таблица 1 – МАТРИЦА ЧАСТОТ

		Классы					Сумма
		1	...	j	...	W	
Значения факторов	1	N_{11}		N_{1j}		N_{1W}	
	...						
	i	N_{i1}		N_{ij}		N_{iW}	$N_i = \sum_{j=1}^{W} N_{ij}$
	...						
	M	N_{M1}		N_{Mj}		N_{MW}	
Суммарное количество признаков				$N_j = \sum_{i=1}^{M} N_{ij}$			$N = \sum_{i=1}^{W} \sum_{j=1}^{M} N_{ij}$
Суммарное количество объектов обучающей выборки				N_j			N

Алгоритм формирования матрицы абсолютных частот.

Объекты обучающей выборки описываются векторами (массивами) $\vec{L} = \{L_i\}$ имеющихся у них признаков:

$$\vec{L} = \{L_i\} = \begin{cases} 1, \textit{ если у объекта есть } \mathbf{i}\textit{-й признак}; \\ 0, \textit{ если у объекта нет } \mathbf{i}\textit{-го признака}. \end{cases}$$

Первоначально в матрице абсолютных частот все значения равны нулю. Затем организуется цикл по объектам обучающей выборки. Если предъявленного объекта относящегося к *j-му* классу есть *i-й* признак, то:

$$N_{ij} = N_{ij} + 1;\ N_i = N_i + 1;\ N_j = N_j + 1;\ N = N + 1$$

Здесь можно провести очень интересную и важную аналогию между способом формирования матрицы абсолютных частот и работой ***многоканальной*** *системы выделения полезного сигнала из шума*. Представим себе, что все объекты, предъявляемые для формирования обобщенного образа некоторого класса в действительности являются различными реализациями одного объекта – "Эйдоса" (в смысле Платона), по-разному зашумленного различными случайными обстоятельствами. И наша задача состоит в том, чтобы подавить этот шум и выделить из него то общее и существенное, что отличает объекты данного класса от объектов других классов. Учитывая, что шум чаще всего является "белым" и имеет свойство при суммировании с самим собой стремиться к нулю, а сигнал при этом наоборот возрастает пропорционально количеству слагаемых, то увеличение объема обучающей выборки приводит ко все лучшему отношению сигнал/шум в матрице абсолютных частот, т.е. к выделению полезной информации из шума. Примерно так мы начинаем постепенно понимать смысл фразы, которую мы сразу не расслышали по телефону и несколько раз переспрашивали. При этом в повторах шум не позволяет понять то одну, то другую часть фразы, но в конце-концов за счет использования памяти и интеллектуальной обработки информации мы понимаем ее всю. Так и *объекты, описанные признаками, можно рассматривать как зашумленные фразы, несущие нам информацию об обобщенных образах классов: "Эйдосах" [12, 13, 14, 15], к которым они относятся. И эту информацию мы выделяем из шума при синтезе модели.*

Для выражения (11):

$$P_{ij} = \frac{N_{ij}}{N_i} \qquad (14)$$

Для выражений (12) и (13):

$$P_{ij} = \frac{N_{ij}}{N_j} \qquad (15)$$

Для выражений (11), (12) и (13):

$$P_i = \frac{N_i}{N};\ P_j = \frac{N_j}{N};$$

$$N_i = \sum_{j=1}^{W} N_{ij};\ N_j = \sum_{i=1}^{M} N_{ij}; \qquad (16)$$

$$N = \sum_{i=1}^{M} N_i = \sum_{j=1}^{W} N_j = \sum_{i=1}^{W} \sum_{j=1}^{M} N_{ij}$$

В (16) использованы обозначения:

N_{ij} – суммарное количество наблюдений в исследуемой выборке *факта*: "действовало *i-е* значение фактора и объект перешел в *j-е* состояние";

N_j – суммарное количество встреч различных факторов у объектов, перешедших в *j-е* состояние;

N_i – суммарное количество встреч *i-го* фактора у всех объектов исследуемой выборки;

N – суммарное количество встреч различных факторов у всех объектов исследуемой выборки.

Формирование матрицы условных и безусловных вероятностей .

На основе анализа матрицы частот (таблица 1) классы можно сравнивать по наблюдаемым частотам признаков только в том случае, если количество объектов по всем классам ***одинаково***, как и ***суммарное количество признаков по классам***. Если же они отличаются, то корректно сравнивать классы можно только по условным и безусловным вероятностям наблюдения признаков, посчитанных на основе матрицы частот (таблица 1) в соответствии с выражениями (14) и (15), в результате чего получается матрица условных и безусловных процентных распределений (таблица 2).

При расчете матрицы условных и безусловных вероятностей N_j из таблицы 1 могут браться либо из предпоследней, либо из последней строки. В 1-м случае N_j представляет собой "Суммарное количество признаков у всех объектов, использованных для формирования обобщенного образа *j-го* класса", а во 2-м случае, это "Суммарное количество объектов обучающей выборки, использованных для формирования обобщенного образа *j-го* класса", соответственно получаем различные, хотя и очень сходные семантические информационные модели, которые мы называем СИМ-1 и СИМ-2. Оба этих вида моделей поддерживаются системой "Эйдос".

Таблица 2 – МАТРИЦА УСЛОВНЫХ И БЕЗУСЛОВНЫХ ВЕРОЯТНОСТЕЙ

		Классы					Безусловная вероятность признака
		1	...	j	...	W	
Значения факторов	1	P_{11}		P_{1j}		P_{1W}	
	...						
	i	P_{i1}		P_{ij}		P_{iW}	P_i
	...						
	M	P_{M1}		P_{Mj}		P_{MW}	
Безусловная вероятность класса				P_j			

Эквивалентность выражений (11) и (12) устанавливается, если подставить в них выражения вероятности P_{ij}, P_j и P_i через частоты наблюдения признаков по классам из (14), (15) и (16). В обоих случаях из выражений (11) и (12) получается *одно и тоже* выражение (17):

$$I_{ij} = Log_2 \frac{N_{ij} N}{N_i N_j} \qquad (17)$$

А из (13) выражение (18), с которым мы и будем далее работать.

$$I_{ij} = Log_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^{\Psi} \qquad (18)$$

При взаимно-однозначном соответствии классов и признаков в *равновероятном детерминистском* случае имеем (таблица 3):

Таблица 3 – МАТРИЦА ЧАСТОТ В РАВНОВЕРОЯТНОМ ДЕТЕРМИНИСТСКОМ СЛУЧАЕ

		Классы					Сумма
		1	**...**	***j***	**...**	***W***	
Значения факторов	***1***	1					1
	...		1				1
	i			1			1
	...				1		1
	M					1	1
Сумма		1	1	1	1	1	***N***

В этом случае к каждому классу относится один объект, имеющий единственный признак. Откуда получаем для всех i и j (19):

$$\forall ij: \; N_{ij} = N_i = N_j = 1 \qquad (19)$$

Таким образом, обобщенная формула А.Харкевича (18) с учетом (19) в этом случае приобретает вид:

$$I_{ij} = Log_2 N^{\Psi} = Log_2 W^{j} \qquad (20)$$

откуда:

$$\boxed{\Psi = \frac{Log_2 W^{j}}{Log_2 N}} \qquad (21)$$

или, учитывая выражение для коэффициента эмерджентности Хартли (8):

$$\Psi = \frac{Log_2 W^{\frac{Log_2 \sum_{m=1}^{M} C_W^m}{Log_2 W}}}{Log_2 N} \qquad (22)$$

Подставив коэффициент эмерджентности А.Харкевича (21) в выражение (18), получим:

$$I_{ij} = Log_2\left(\frac{N_{ij}N}{N_i N_j}\right)^{\Psi} = Log_2\left(\frac{N_{ij}N}{N_i N_j}\right)^{\frac{Log_2 W^j}{Log_2 N}} =$$

$$= \frac{Log_2 W^j}{Log_2 N}\left(Log_2\left(\frac{N_{ij}}{N_i N_j}\right) + Log_2 N\right) =$$

$$= Log_2\left(\frac{N_{ij}}{N_i N_j}\right)^{\frac{Log_2 W^j}{Log_2 N}} + Log_2 W^j$$

или окончательно:

$$\boxed{I_{ij} = Log_2\left(\frac{N_{ij}}{N_i N_j}\right)^{\frac{Log_2 W^j}{Log_2 N}} + Log_2 W^j} \qquad (23)$$

Отметим, что 1-я задача получения системного обобщения формул Хартли и Харкевича и 2-я задача получения такого обобщения формулы Харкевича, которая удовлетворяет принципу соответствия с формулой Хартли – это две разные задачи. 1-я задача является более общей и при ее решении, которое приведено выше, *автоматически* решается и 2-я задача, которая является, таким образом, частным случаем 1-й.

Однако, представляет самостоятельный интерес и частный случай, в результате которого получается формула Харкевича удовлетворяющая в *равновероятном детерминистском* случае принципу соответствия с классической формулой Хартли (1), а не с ее системным обобщением (2) и (3). Ясно, что эта формула получается из (23) при $j=1$.

$$I_{ij} = Log_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{Log_2 W}{Log_2 N}} + Log_2 W \qquad (24)$$

Из выражений (21) и (22) видно, что в этом частном случае, т.е. когда система эквивалентна множеству (*M=1*), коэффициент эмерджентности А.Харкевича приобретает вид:

$$\Psi = \frac{Log_2 W}{Log_2 N} \qquad (25)$$

На практике для численных расчетов на ***удобнее пользоваться*** не выражениями (23) или (24), а ***формулой (26)***, которая получается непосредственно из (18) после подстановки в него выражения (25):

$$\boxed{I_{ij} = \frac{Log_2 W}{Log_2 N} \times Log_2 \frac{N_{ij} N}{N_i N_j}} \qquad (26)$$

В классическом анализе Шеннона идет речь лишь о передаче символов по одному информационному каналу от одного источника к одному приемнику. Его интересует прежде всего передача самого сообщения.

В данной статье ставится другая задача: идентифицировать или распознать информационный источник по сообщению от него. Поэтому метод Шеннона был обобщен путем учета в математической модели возможности существования многих источников информации, о которых к приемнику по зашумленному каналу связи приходят не отдельные символы-признаки, а сообщения, состоящие из последовательностей символов (признаков) любой длины.

Следовательно, ставится задача идентификации информационного источника по сообщению от него, полученному приемником по зашумленному каналу. Метод, являющийся обобщением метода К.Шеннона, позволяет применить классическую теорию информации для построения моделей систем распознавания образов и принятия решений, ориентированных на применение для синтеза адаптивных АСУ сложными объектами.

Для решения поставленной задачи необходимо вычислять не средние информационные характеристики, как в теории Шеннона, а количество информации, содержащееся в конкретном i-м признаке (символе) о том, что он пришел от данного j-го источника информации. Это позволит определить и суммарное количество информации в сообщении о каждом информационном источнике, что дает интегральный критерий для идентификации или прогнозирования состояния объекта.

Логично предположить, что среднее количество информации, содержащейся в системе признаков о системе классов

$$I(Y,X)=\sum_{j=1}^{W}\sum_{i=1}^{M}p_{ij}Log_2\frac{p_{ij}}{p_i p_j}, \qquad (27)$$

является ничем иным, как усреднением (с учетом условной вероятности наблюдения) "индивидуальных количеств информации", которые содержатся в конкретных признаках о конкретных классах (источниках), т.е.:

$$i(x_j,y_i)=Log_2\frac{p_{ij}}{p_i p_j}. \qquad (28)$$

Это выражение определяет так называемую "плотность информации", т.е. количество информации, которое содержится в одном отдельно взятом факте наблюдения *i-го* символа (признака) на приемнике о том, что этот символ (признак) послан *j-м* источником.

Если в сообщении содержится *M* символов, то суммарное количество информации о принадлежности данного сообщения *j-му* информационному источнику (классу) составляет:

$$i(x_j)=\sum_{i=1}^{M}Log_2\frac{p_{ij}}{p_i p_j}. \qquad (29)$$

Необходимо отметить, что применение сложения в выражении (29) является вполне корректным и оправданным, так как информация с самого начала вводилась как аддитивная величина, для которой операция сложения является корректной.

Преобразуем выражение (29) к виду, более удобному для применения на практике для численных расчетов. Для этого *традиционным для теории информации Шеннона способом* выразим вероятности встреч признаков через частоты их наблюдения:

$$P_{ij}=\frac{1}{N_{ij}};\ P_i=\frac{1}{N_i};\ P_j=\frac{1}{N_j}. \qquad (30)$$

Подставив (30) в (29), получим:

$$i(x_j)=\sum_{i=1}^{M}Log_2\frac{N_{ij}}{N_i N_j}. \qquad (31)$$

Если ранжировать классы в порядке убывания суммарного количества информации о принадлежности к ним, содержащейся в данном сообщении (т.е. описании объекта), и выбирать первый из них, т.е. *тот, о котором в сообщении содержится наибольшее количество информации*, то мы получим обоснованную статистическую процедуру, основанную на классической теории информации, оптимальность которой доказывается в фундаментальной ***лемме Неймана-Пирсона*** [22].

Подставим значения вероятностей из (30) в (28) и получим выражением для плотности информации Шеннона, выраженное не через вероятности, а через частоты наблюдения символов, которые рассматриваются как признаки объектов, т.е. количество информации, содержащееся в отдельном ***i****-м* признаке о том, что другом конце канала связи находится ***j****-й* объект (32):

$$i(x_j, y_i) = Log_2 \frac{N_{ij}}{N_i N_j}. \qquad (32)$$

Сравнивая выражения (23) и (32) видим, что в системном обобщении формулы Харкевича 1-е слагаемое ***практически тождественно*** *выражению Шеннона для плотности информации, а* 2-е слагаемое представляющем собой *плотность информации по Хартли.*

Различия состоят в том, что в выражении (23) это слагаемое возведено в степень, имеющую смысл коэффициента эмерджентности Харкевича. Поэтому вполне оправданным называть это слагаемое не коэффициентом эмерджентности Харкевича, а коэффициентом эмерджентности Шеннона-Харкевича. Необходимо отметить также, что значения частот в этих формулах связаны с вероятностями несколько *различным* образом (выражения 14-16 и 30).

Из этого следует также, что полученное выражение (23) представляет собой ***нелинейную суперпозицию*** выражений для плотности информации Шеннона и Хартли, и, таким образом, является ***обобщающим*** выражением для плотности информации, которое при различных условиях асимптотически переходит в классические выражения Хартли и Харкевича, а от выражения Шеннона отличается лишь константой, т.е. 2-м слагаемым, характеризующим мощность множества состояний объекта в модели.

Это позволяет нам обоснованно высказать гипотезу о том, что системная теория информации (СТИ), базирующаяся на выражении (23) для плотности информации, является более общей, чем теории Хартли, Шеннона и Харкевича и асимптотически связана с ними через принцип соответствия (рисунок 1).

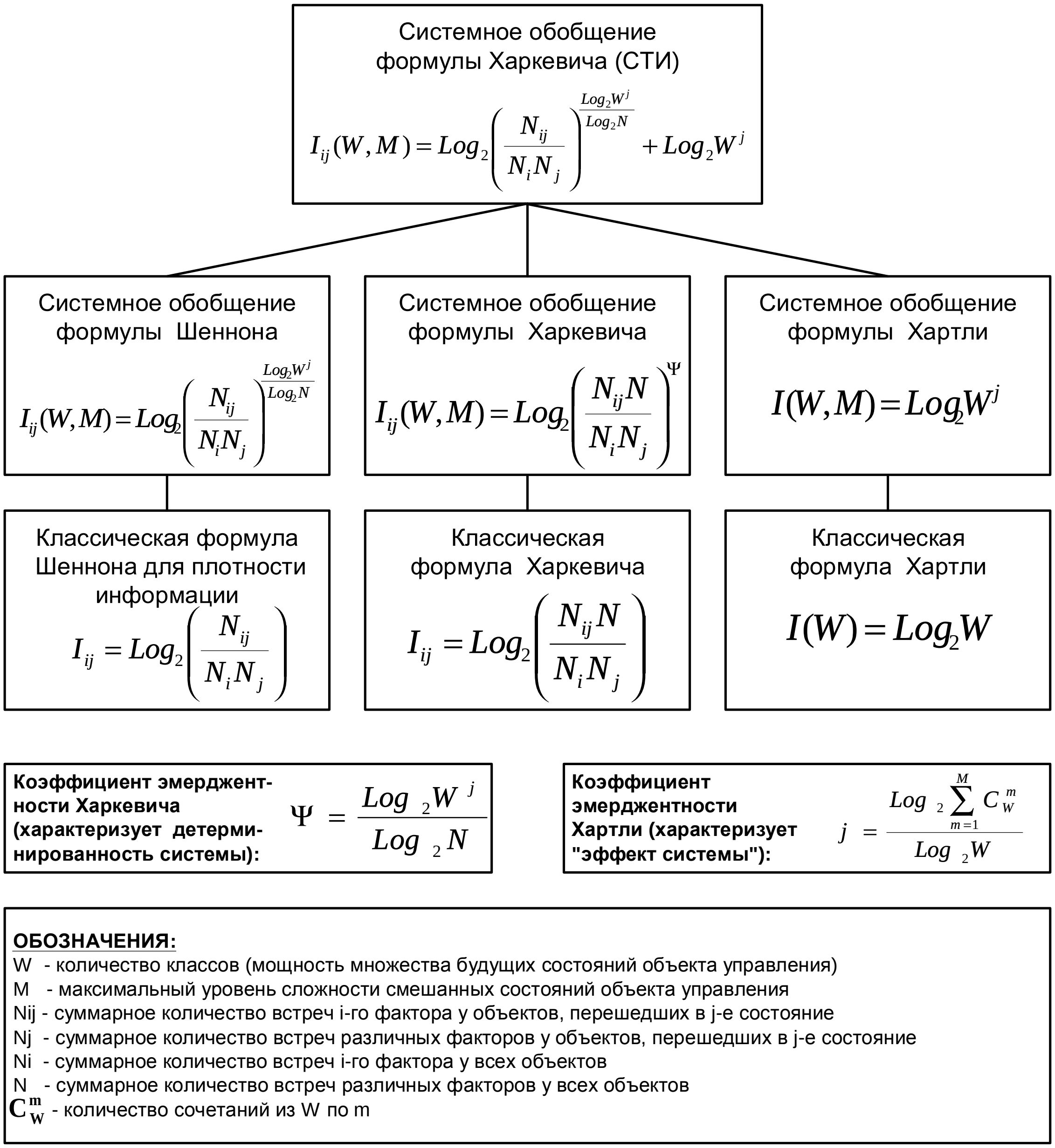

Рисунок 1. Генезис системной (эмерджентной) теории информации

Взаимосвязь системной меры целесообразности информации со статистикой c^2 *и новая мера уровня системности предметной области*

Статистика χ^2 представляет собой сумму вероятностей совместного наблюдения признаков и объектов по всей корреляционной матрице или определенным ее подматрицам (т.е. *сумму относительных **отклонений** частот совместного наблюдения признаков и объектов **от среднего***):

$$c^2 = \sum_{j=1}^{W} \sum_{i=1}^{M} \frac{(N_{ij} - t)^2}{t}, \qquad (33)$$

где

N_{ij} – фактическое количество встреч i-го признака у объектов j-го класса;

t – ожидаемое количество встреч i-го признака у объектов j-го класса (34).

$$t = \frac{N_i N_j}{N}. \qquad (34)$$

Отметим, что статистика χ^2 математически связана с количеством информации в системе признаков о классе распознавания в соответствии с системным обобщением формулы Харкевича для плотности информации (18):

$$I_{ij} = Log_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^{\Psi}, \qquad (35)$$

а именно, из (34) и (35) получаем:

$$I_{ij} = Log_2 \left(\frac{N_{ij}}{t} \right)^{\Psi}. \qquad (36)$$

Из (36) очевидно:

$$I_{ij} = \Psi(Log_2 N_{ij} - Log_2 t). \qquad (37)$$

Сравнивая выражения (33) и (37), видим, что числитель в выражении (33) под знаком суммы отличается от выражения (37) только тем, что в выражении (37) вместо значений N_{ij} и t взяты их логарифмы. Так как логарифм является *монотонно возрастающей* функцией аргумента, то *введение логарифма* ***не меняет*** *общего характера поведения функции*.

Фактически это означает, что:

$$\begin{cases} если \;\; N_{ij} < t \; то \;\; c_{ij} > 0, \, I_{ij} < 0 \\ если \;\; N_{ij} = t \; то \;\; c_{ij} = 0, \, I_{ij} = 0 \\ если \;\; N_{ij} > t \; то \;\; c_{ij} > 0, \, I_{ij} > 0 \end{cases}. \qquad (38)$$

<u>Из изложенного следует интерпретация системной меры информации (35) с учетом статистики χ^2 (33):</u> *если фактическая вероятность наблюдения **i**-го признака при предъявлении объекта **j**-го класса равна ожидаемой (средней), то наблюдение этого признака* ***не несет никакой информации о принадлежности объекта к данному классу.*** *Если она выше средней, то это свидетельствует в пользу того, что предъявлен объект данного класса, если ниже – то другого.*

Поэтому наличие статистической связи (информации) между признаками и классами распознавания, т.е. отличие вероятностей их совместных наблюдений от предсказываемого в соответствии со случайным нормальным распределением, приводит к увеличению фактической статистики χ^2 по сравнению с теоретической величиной.

Из этого следует возможность использования в качестве количественной меры степени выраженности закономерностей в предметной области не матрицы абсолютных частот и меры χ^2, а новой меры H, основанной на матрице информативностей и системном обобщении формулы Харкевича для количества информации:

$$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^{W} \sum_{i=1}^{M} \left(I_{ij} - \bar{I}\right)^2}\ , \qquad (39)$$

где

$\bar{I} = \frac{1}{W \cdot M} \sum_{j=1}^{W} \sum_{i=1}^{M} I_{ij}$ – средняя информативность признаков по матрице информативностей.

Меру H в выражении (39) предлагается назвать обобщенным критерием степени сформированности модели Харкевича.

Значение данной меры показывает среднее отличие количества информации в факторах о будущих состояниях активного объекта управления от среднего количества информации в факторе (которое при больших выборках близко к 0). По своей математической форме эта мера сходна с мерами для значимости (интегральной информативности) факторов и степени сформированности образов классов и *коррелирует с* ***объемом*** *неортонормированного семантического информационного пространства классов и семантического информационного пространства атрибутов*.

Вышеописанная математическая модель обеспечивает ***инвариантность*** *результатов ее синтеза* относительно следующих параметров обучающей выборки: *суммарное количество и порядок ввода анкет обучающей выборки;* количество анкет обучающей выборки по каждому классу распознавания; суммарное количество признаков во всех анкетах обучающей выборки; суммарное количество признаков по классам распознавания; количество признаков и их порядок в отдельных анкетах обучающей выборки. Это обеспечивает высокую степень качества решения задач распознавания на неполных и разнородных (в вышеперечисленных аспектах) данных как обучающей, так и распознаваемой выборки, т.е. при таких статистических характеристиках потоков этих данных, которые чаще всего и встречается на практике и которыми невозможно или очень сложно управлять.

Получение матрицы знаний (информативностей).

На основе анализа матрицы условных и безусловных вероятностей (таблица 3) наблюдений признаков по классам и всей выборке можно сравнивать признаки друг с другом по их роли для сравнения классов друг с другом и конкретных объектов с обобщенными классами. При этом существует 3 основных группы признаков:

Группа 1-я. Которые в одном классе встречаются, а в других нет. Это *детерминистские* признаки, обнаружение такого признака у объекта однозначно определяет его принадлежность к соответствующему классу.

***Группа 2-я*.** Которые в одном классе встречаются чаще, чем в других. Это *статистические* признаки, обнаружение такого признака у объекта несет некоторую информацию о его принадлежности к соответствующему классу.

***Группа 3-я*.** Которые в разных классах встречаются одной и той же вероятностью. Это признаки, обнаружение которых у объекта не несет никакой информации о его принадлежности к тем или иным классам.

Таким образом мы видим, что если используя таблицу 3 анализировать условные вероятности (или процентные распределения) признаков по классам, то можно вынести правдоподобные суждения о принадлежности объектов, обладающих этими признакам к тем или иным классам.

Однако в таком методе сравнения есть по крайней мере два ***существенных недостатка:***

1. Для того, чтобы отнести признак к одной из вышеперечисленных групп необходимо ***сравнивать*** вероятности его наблюдения по классам, т.е. каждый раз при таком сравнении выполнять соответствующую необходимую для этого работу.

2. При отнесении признака ко 2-й группе этого самого по себе еще недостаточно для его использования с целью идентификации объекта, а необходимо еще оценить ***количество*** информации, которое содержится в факте обнаружения у объекта этого признака о принадлежности этого объекта к каждому из классов, а для этого *необходим соответствующий математический и численный **метод**.*

Что касается 1-го недостатка, то о нем можно сказать, что для реальных задач большой размерности выполнение этого сравнения вручную практически невозможно, а значит тем более невозможно и использование результатов этого сравнения для решения задач идентификации, прогнозирования и поддержки принятия решений, а тем более для исследования предметной области путем исследования ее модели. Все это обусловлено тем, что результат сравнения вероятностей встречи признака по классам не представляется при ручной обработке в количественной форме некоторого ***одного*** числа: ***частного критерия***, величина и знак которого отражали бы результат такого сравнения.

2-й недостаток преодолевается методом, который предложен А.Харкевичем в выражениях (11) и (12) и уточнен нами в системном обобщении этих выражений (18). *В этом методе предложено сравнивать не условные вероятности наблюдения признаков по различным классам* *друг с другом, а условную вероятность наблюдения признака по классу с безусловной вероятностью его наблюдения по всей выборке.*

Это предложение по своей сути полностью соответствует известному статистическому *методу отклонений от средних* и ***нормативному подходу***, когда в качестве *базы сравнения* выбирается *норма*, т.е. *среднее* по всей группе. На основе этого подхода формируются и критерии сравнения, т.е. можно сказать, что ***критериальный подход*** изначально основан на нормативном.

Если такое сравнение провести по всем признакам и классам, то получится матрица, снимающая оба указанных недостатка: используя выражение (18) и данные таблицы 1 непосредственно прямым счетом получаем *матрицу знаний* (таблица 4):

Таблица 4 – МАТРИЦА ЗНАНИЙ (ИНФОРМАТИВНОСТЕЙ)

		Классы					Значимость фактора
		1	...	j	...	W	
Значения факторов	1	I_{11}		I_{1j}		I_{1W}	$s_1 = \sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{1j}-\bar{I}_1\right)^2}$
	...						
	i	I_{i1}		I_{ij}		I_{iW}	$s_i = \sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{ij}-\bar{I}_i\right)^2}$
	...						
	M	I_{M1}		I_{Mj}		I_{MW}	$s_M = \sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{Mj}-\bar{I}_M\right)^2}$
Степень редукции класса		s_1		s_j		s_W	$H = \sqrt[2]{\frac{1}{(W\cdot M-1)}\sum_{j=1}^{W}\sum_{i=1}^{M}\left(I_{ij}-\bar{I}\right)^2}$

Здесь – $\bar{I}_i$ это *среднее* количество знаний в *i-м* значении фактора:

$$\bar{I}_i = \frac{1}{W}\sum_{j=1}^{W} I_{ij}$$

При расчете матрицы знаний N_j из таблицы 1 могут браться либо из предпоследней, либо из последней строки. В 1-м случае N_j представляет собой "Суммарное количество признаков у всех объектов, использованных для формирования обобщенного образа *j-го* класса", а во 2-м случае, это "Суммарное количество объектов обучающей выборки, использованных для формирования обобщенного образа *j-го* класса", соответственно получаем различные, хотя и очень сходные семантические информационные модели, которые мы называем СИМ-1 и СИМ-2. Оба этих вида моделей поддерживаются системой "Эйдос".

Количественные значения коэффициентов I_{ij} таблицы 4 являются знаниями о том, что "объект перейдет в j-е состояние" если "на объект действует i-е значение фактора".

Принципиально важно, что эти весовые коэффициенты не определяются экспертами на основе опыта интуитивным неформализуемым способом, а *рассчитываются непосредственно на основе эмпирических данных на основе теоретически обоснованной модели, хорошо зарекомендовавшей себя на практике при решении широкого круга задач в различных предметных областях*.

Когда количество информации $I_{ij}>0$ – ***i–й*** фактор способствует переходу объекта управления в ***j-е*** состояние, когда $I_{ij}<0$ – препятствует этому переходу, когда же $I_{ij}=0$ – никак не влияет на это. В векторе ***i-го*** фактора (строка матрицы информативностей) отображается, какое количество информации о переходе объекта управления в каждое из будущих состояний содержится в том факте, что данный фактор действует. В векторе ***j-го*** состояния класса (столбец матрицы информативностей) отображается, какое количество информации о переходе объекта управления в соответствующее состояние содержится в каждом из факторов.

Таким образом, матрица информативностей (таблица 4) является обобщенной таблицей решений, в которой входы (факторы) и выходы (будущие состояния объекта управления) связаны друг с другом не с помощью классических (Аристотелевских) импликаций, принимающих только значения: "Истина" и "Ложь", а ***различными значениями истинности, выраженными в битах*** *и принимающими значения от положительного теоретически-максимально-возможного ("Максимальная степень истинности"), до теоретически неограниченного отрицательного ("Степень ложности").*

Фактически предложенная модель позволяет осуществить *синтез обобщенных таблиц решений* для различных предметных областей непосредственно на основе эмпирических исходных данных и *продуцировать на их основе прямые и обратные правдоподобные (нечеткие) логические рассуждения по неклассическим схемам* ***с различными расчетными значениями истинности***, являющимся обобщением классических импликаций.

Таким образом данная *модель позволяет рассчитать какое* ***количество информации*** *содержится в любом факте о наступлении любого события в любой предметной области, причем для этого не требуется повторности этих фактов и событий*. Если же эти повторности осуществляются и при этом наблюдается некоторая вариабельность значений факторов, обуславливающих наступление тех или иных событий, то модель обеспечивает ***многопараметрическую типизацию***, т.е. синтез обобщенных образов классов или категорий наступающих событий с количественной оценкой степени и знака влияния на их наступление различных значений факторов. Причем эти значения факторов могут быть как количественными, так и качественными и измеряться в любых единицах измерения, в любом случае в модели оценивается количество информации которое в них содержится о наступлении событий, переходе объекта управления в определенные состояния или просто о его принадлежности к тем или иным классам.

Данная модель позволяет прогнозировать поведение объекта управления при воздействии на него не только одного, но и целой ***системы*** *факторов*:

$$I_j = f(\vec{I}_{ij}). \qquad (40)$$

В теории принятия решений скалярная функция I_j векторного аргумента называется ***интегральным критерием***. Основная проблема состоит в выборе такого аналитического *вида функции* интегрального критерия, который обеспечил бы эффективное решение сформулированной выше задачи АСУ.

Учитывая, что ***частные критерии*** (18) имеют смысл количества информации, а информация по определению является *аддитивной* функцией, предлагается ввести *интегральный критерий*, как аддитивную функцию от частных критериев в виде:

$$I_j = (\vec{I}_{ij}, \vec{L}_i). \qquad (41)$$

В выражении (41) круглыми скобками обозначено скалярное произведение, т.е. свертка. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^{M} I_{ij} L_i, \qquad (42)$$

где:

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор *j-го* класса-состояния объекта управления;

$\vec{L}_i = \{L_i\}$ – вектор состояния предметной области, включающий все виды факторов, характеризующих объект управления, возможные управляющие воздействия и окружающую среду (массив-локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, \textit{ если } i\textit{-й фактор действует;} \\ a_i, \textit{ где}: 0 < a_i < 1, \textit{ если } i\textit{-й фактор действует с истинностью } a_i \\ 0, \textit{ если } i\textit{-й фактор не действует.} \end{cases}$$

В реализованной модели значения координат вектора состояния предметной области принимались равными либо 1 (фактор действует), либо 0 (фактор не действует).

Таким образом, *интегральный критерий представляет собой суммарное количество информации, содержащееся в системе значений факторов различной природы (т.е. факторах, характеризующих объект управления, управляющее воздействие и окружающую среду) о переходе объекта управления в то или иное будущее состояние.*

В многокритериальной постановке задача прогнозирования состояния объекта управления, при оказании на него заданного многофакторного управляющего воздействия I_j, сводится к максимизации интегрального критерия:

$$j^* = \arg\max_{j \in J}((\vec{I}_{ij}, \vec{L}_i)), \qquad (43)$$

т.е. к выбору такого состояния объекта управления, для которого интегральный критерий максимален.

Задача принятия решения о выборе наиболее эффективного управляющего воздействия является *обратной задачей* по отношению к задаче максимизации интегрального критерия (идентификации и прогнозирования), т.е. вместо того, чтобы по набору факторов прогнозировать будущее состояние объекта, наоборот, по заданному (целевому) состоянию объекта определяется такой набор факторов, который с наибольшей эффективностью перевел бы объект управления в это состояние.

Предлагается еще одно обобщение фундаментальной леммы Неймана-Пирсона, основанное на косвенном учете корреляций между информативностями в векторе состояний при использовании средних по векторам. Соответственно, вместо простой суммы количеств информации предлагается использовать корреляцию между векторами состояния и объекта управления, которая количественно измеряет степень сходства этих векторов:

$$I_j = \frac{1}{s_j s_l A} \sum_{i=1}^{M} \left(I_{ij} - \bar{I}_j\right)\left(L_i - \bar{L}\right), \qquad (44)$$

где:

$\bar{I}_j$ – средняя информативность по вектору класса;

$\bar{L}$ – среднее по вектору идентифицируемой ситуации (объекта).

s_j – среднеквадратичное отклонение информативностей вектора класса;

s_l – среднеквадратичное отклонение по вектору распознаваемого объекта.

Выражение (44) получается непосредственно из (42) после замены координат перемножаемых векторов их стандартизированными значениями:

$$I_{ij} \to \frac{I_{ij} - \bar{I}_j}{s_j}, \quad L_i \to \frac{L_i - \bar{L}}{s_l}.$$

Необходимо отметить, что выражение для интегрального критерия сходства (42) по своей математической форме является ***корреляцией*** двух векторов. Это означает, что если эти вектора являются суммой двух сигналов: полезного и белого шума, то *при расчете интегрального критерия белый шум практически не будет играть никакой роли*, т.е. его корреляция с самими собой равна нулю по определению. Поэтому интегральный критерий сходства объекта со случным набором признаков с любыми образами классов, или реального объекта с образами классов, сформированными случайным образом, будет равен нулю. Это означает, что ***выбранный интегральный критерий сходства является высокоэффективным средством подавления белого шума и выделения полезной информации из шума***, который неизбежно присутствует в эмпирических данных.

Важно также отметить ***неметрическую природу*** предложенного интегрального критерия сходства, благодаря чему *его применение является* ***корректным*** *и при* ***неортонормированном*** *семантическом информационном пространстве, каким оно в подавляющем количестве случае и является, т.е. в общем случае.*

Результат прогнозирования поведения объекта управления, описанного данной системой факторов, представляет собой список его возможных будущих состояний, в котором они расположены в порядке убывания суммарного количества информации о переходе объекта управления в каждое из них.

Сравнение, идентификация и прогнозирование как разложение векторов объектов в ряд по векторам классов (объектный анализ)

Выше были введены неметрические интегральные критерии сходства объекта, описанного массивом-локатором L_i с обобщенными образами классов I_{ij} (выражения 40-42).

Для непрерывного случая выражение (42) принимает вид:

$$I_j = \int_1^M L(i) I_j(i) di \,. \qquad (45)$$

Таким образом, выражение (45) представляет собой *обобщение* интегрального критерия сходства конкретного объекта и обобщенного класса (42) для *непрерывного* случая в координатной форме.

Отметим, что коэффициенты ряда Фурье **(24)** *по своей математической форме и смыслу* *<u>сходны</u>* *с* ***ненормированными*** *коэффициентами корреляции, т.е. по сути скалярными произведениями для непрерывных функций в координатной форме: выражением* **(45)** *между разлагаемой в ряд кривой* **f(x)** *и функциями* **Sin** *и* **Cos** *различных частот и амплитуд* [1].

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}\left(a_n\cos\left(\frac{npx}{L}\right)+b_n\sin\left(\frac{npx}{L}\right)\right)$$

где :

$$a_0=\frac{1}{L}\int_{-L}^{L}f(x)dx \qquad , \qquad (46)$$

$$a_n=\frac{1}{L}\int_{-L}^{L}f(x)\cos\left(\frac{npx}{L}\right)dx$$

$$b_n=\frac{1}{L}\int_{-L}^{L}f(x)\sin\left(\frac{npx}{L}\right)dx$$

где n={1, 2, 3,…} – натуральное число.

Из сравнения выражений (45) и (46) следует вывод о том, что ***процесс идентификации и прогнозирования (распознавания), реализованный в предложенной математической модели, может рассматриваться <u>как разложение вектора-локатора распознаваемого объекта в ряд по векторам информативностей классов распознавания</u> (которые представляют собой <u>произвольные</u> функции, сформированные при синтезе модели путем многопараметрической типизации на основе эмпирических данных).***

Например, результаты идентификации представим на рисунке 2:

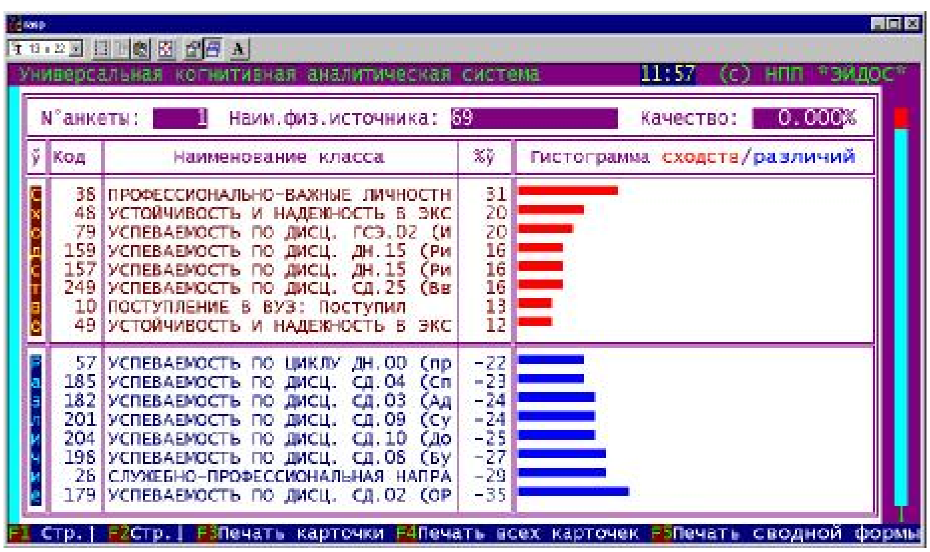

Рисунок 2. Пример разложения профиля курсанта усл.№69 в ряд по обобщенным образам классов

Продолжая развивать аналогию с разложением в ряд, данный результат идентификации можно представить в векторной *аналитической* форме:

$$\begin{aligned}\vec{K}_{усл.№} &= 0{,}31\cdot\vec{I}(38)+0{,}20\cdot\vec{I}(48)+0{,}20\cdot\vec{I}(79)+0{,}16\cdot\vec{I}(159)+\\ &+0{,}16\cdot\vec{I}(157)+0{,}16\cdot\vec{I}(249)+\ldots-0{,}35\cdot\vec{I}(179)-0{,}29\cdot\vec{I}(26)-\\ &-0{,}27\cdot\vec{I}(198)-0{,}25\cdot\vec{I}(204)-0{,}24\cdot\vec{I}(201)-0{,}24\cdot\vec{I}(182)-\ldots\end{aligned}$$

или в координатной форме, более удобной для численных расчетов:

$$\boxed{K(i)=\sum_{j=1}^{W}\left(I(j)\cdot I(i,j)\right)}, \qquad (47)$$

Предполагается, что $\vec{L}_i=\{L_i\}\approx K(i)$. *Таким образом массив-локатор, характеризующий распознаваемый объект, рассматривается как сумма произведений профилей классов на интегральный критерий сходства массива-локатора с этими профилями (т.е. взвешенная суперпозиция или разложение в ряд по профилям классов).*

В выражении (47):

I(j) – интегральный критерий сходства массива-локатора, описывающего состояние объекта и *j*-го класса, рассчитываемый, согласно выражений (42) или (44):

$$I_j=\frac{1}{\sigma_j\sigma_l M}\sum_{i=1}^{M}\left(I_{ij}-\bar{I}_j\right)\left(L_i-\bar{L}\right), \qquad (48)$$

I(i,j) – вектор обобщенного образа *j*-го класса, координаты которого рассчитываются в соответствии с системным обобщением формулы Харкевича (18):

$$I_{ij}=Log_2\left(\frac{N_{ij}}{N_iN_j}\right)^{\frac{Log_2W^j}{Log_2N}}+Log_2W^j \qquad (49)$$

<u>Примечание:</u> обозначения *I(i,j)* и I_{ij}, и т.п. эквивалентны. Смысл всех переменных, входящих в выражения (48) и (49) раскрыт выше.

При дальнейшем развитии данной аналогии естественно возникают вопросы: о полноте, избыточности и ортонормированности системы векторов классов как функций, по которым проводится разложение вектора объекта; о сходимости, т.е. вообще возможности и корректности такого разложения.

В общем случае вектор объекта совершенно не обязательно должен разлагаться в ряд по векторам классов таким образом, что сумма ряда во всех точках точно совпадала со значениями исходной функции. Это означает, что система векторов классов может быть ***неполна*** по отношению к профилю распознаваемого объекта, и, тем более, всех возможных объектов.

Предлагается считать не разлагаемые в ряд, т.е. плохо распознаваемые объекты, суперпозицией хорошо распознаваемых объектов ("похожих" на те, которые использовались для формирования обобщенных образов классов), и объектов, которые и не должны распознаваться, так как объекты этого типа <u>не встречались в обучающей выборке и не использовались</u> для формирования обобщенных образов классов, а также не относятся к представляемой обучающей выборкой генеральной совокупности.

Нераспознаваемую компоненту можно рассматривать либо как шум, либо считать ее полезным сигналом, несущим ценную информацию о неисследованных объектах интересующей нас предметной области (в зависимости от целей и тезауруса исследователей). Первый вариант не приводит к осложнениям, так как примененный в математической модели алгоритм сравнения векторов объектов и классов, основанный на вычислении нормированной корреляции Пирсона (сумма произведений), является *весьма устойчивым к наличию белого шума* в идентифицируемом сигнале. Во втором варианте необходимо дообучить систему распознаванию объектов, несущих такую компоненту (в этой возможности и заключается адаптивность модели). Технически этот вопрос решается просто копированием описаний плохо распознавшихся объектов из распознаваемой выборки в обучающую, их идентификацией экспертами и дообучением системы. Кроме того, может быть целесообразным расширить справочник классов распознавания новыми классами, соответствующими этим объектам, и осуществить пересинтез модели.

Однако на практике гораздо чаще наблюдается противоположная ситуация (можно даже сказать, что она типична), когда система векторов ***избыточна***, т.е. в системе классов распознавания есть очень похожие классы (между которыми имеет место высокая корреляция, наблюдаемая в режиме: "кластерно-конструктивный анализ"). Практически это означает, что в системе сформировано несколько практически одинаковых образов с разными наименованиями. Для исследователя это само по себе является очень ценной информацией. Однако если исходить только из потребности разложения распознаваемого объекта в ряд по векторам классов (чтобы определить суперпозицией каких образов он является, т.е. "разложить его на компоненты"), то наличие сильно коррелирующих друг с другом векторов представляется неоправданным, так как просто увеличивает размерности данных, внося в них мало нового по существу. Поэтому возникает задача *исключения избыточности системы классов распознавания*, т.е. выбора из всей системы классов распознавания такого минимального их набора, в котором профили классов минимально коррелируют друг с другом, *т.е. ортогональны в фазовом пространстве признаков*. Это условие в теории рядов называется "ортонормируемостью" системы базовых функций, а в факторном анализе связано с идеей выделения "главных компонент".

В предлагаемой математической модели реализованы два варианта выхода из данной ситуации:

1) исключение неформирующихся, расплывчатых классов;

2) объединение почти идентичных по содержанию (дублирующих друг друга) классов.

Однако выбрать нужный вариант и реализовать его, используя соответствующие режимы, пользователь технологии АСК-анализа должен сам. Вся необходимая и достаточная информация для принятия соответствующих решений предоставляется пользователю инструментария АСК-анализа.

Если считать, что функции образов составляют формально-логическую систему, к которой применима теорема Геделя, то можно сформулировать эту теорему для данного случая следующим образом: "Для любой системы базисных функций в принципе всегда может существовать по крайней мере одна такая функция, что она не может быть разложена в ряд по данной системе базисных функций, т.е. ***функция, которая является ортонормированной ко всей системе базисных функций в целом"***. Поэтому для адекватного отражения подобных функций в модели необходимо повышение размерности семантического информационного пространства.

Очевидно, не взаимосвязанными друг с другом могут быть только четко оформленные, детерминистские образы, т.е. образы с высокой степенью редукции ("степень сформированности конструкта"). Поэтому в процессе выявления взаимно-ортогональных базисных образов, в первую очередь, будут выброшены аморфные "расплывчатые" образы, которые связаны практически со всеми остальными образами.

В некоторых случаях результат такого процесса представляет интерес, и это делает оправданным его реализацию. Однако можно предположить, что наличие расплывчатых образов в системе является оправданным, так как в этом случае система образов не будет формальной и подчиняющейся теореме Геделя. Следовательно, система распознавания будет более полна в том смысле, что увеличится вероятность идентификации *любого объекта*, предъявленного ей на распознавание. Конечно, уровень сходства с аморфным образом не может быть столь высоким, как с четко оформленным. Поэтому в этом случае более уместно применить термин "ассоциация" или нечеткая, расплывчатая идентификация, чем "однозначная идентификация".

Итак, можно сделать следующий вывод: допустимость в математической модели СК-анализа не только четко оформленных (детерминистских) образов, но и образов аморфных, нечетких, расплывчатых является важным достоинством данной модели. Это обусловлено тем, что данная модель обеспечивает корректные результаты анализа, идентификации и прогнозирования даже в тех случаях, когда модели идентификации и информационно-поисковые системы детерминистского типа традиционных АСУ практически неработоспособны. В этих условиях данная модель СК-анализа работает как система *ассоциативной (нечеткой) идентификации*.

Таким образом, в предложенной семантической информационной модели при идентификации и прогнозировании, по сути, осуществляется разложение векторов идентифицируемых объектов по векторам классов распознавания, т.е. осуществляется ***"объектный анализ"*** *(по аналогии с спектральным, гармоническим или* ***Фурье-анализом****),* ***что позволяет рассматривать идентифицируемые объекты как суперпозицию обобщенных образов классов различного типа с различными амплитудами (25). При этом вектора обобщенных образов классов, с математической точки зрения, представляют собой произвольные функции и не обязательно образуют полную и не избыточную (ортонормированную) систему функций.***

Для любого объекта всегда существует такая система базисных функций, что вектор объекта может быть представлен в форме линейной суперпозиции (суммы) этих базисных функций с различными амплитудами. Это утверждение, по-видимому, является одним из следствий фундаментальной теоремы А.Н. Колмогорова, доказанной им в 1957 году (О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения // Докл. АН СССР, Т. 114, С. 953–956, 1957).

Теорема Колмогорова: Любая непрерывная функция от n переменных $F(x_1, x_2, ..., x_n)$ может быть представлена в виде:

$$F(x_{1,}, x_2, ..., x_n) = \sum_{j=1}^{2n+1} \left(g_j \sum_{i=1}^{n} \left(h_{ij}(x_i) \right) \right),$$

где g_j и h_{ij} – непрерывные функции, причем h_{ij} не зависят от функции F.

Эта теорема означает, что для реализации функций многих переменных достаточно операций суммирования и композиции функций одной переменной. Удивительно, что в этом представлении лишь функции g_j зависят от представляемой функции F, а функции h_{ij} универсальны. ***Необходимо отметить, что теорема Колмогорова является обобщением теоремы В.И. Арнольда (1957), которая дает решение 13-й проблемы Гильберта.***

К сожалению, определение вида функций h_{ij} и g_j для данной функции F представляет собой математическую проблему, для которой пока не найдено общего строгого решения.

В данной работе предлагается рассматривать предлагаемую семантическую информационную модель как один из вариантов решения этой проблемы. В этом контексте функция F интерпретируется как образ идентифицируемого объекта, функция h_{ij} – образ j-го класса, а функция g_j – мера сходства образа объекта с образом класса.

Таким образом В статье кратко описана математическая сущность предложенной автором системной теории информации СТИ), являющейся математической моделью системно-когнитивного анализа (СК-анализ) и реализуемой в его программном инструментарии – универсальной аналитической системе "Эйдос" [20, 21].

Литература[1]

1. Lutsenko E.V. Conceptual principles of the system (emergent) information theory & its application for the cognitive modelling of the active objects (entities). 2002 IEEE International Conference on Artificial Intelligence System (ICAIS 2002). –Computer society, IEEE, Los Alamos, California, Washington-Brussels-Tokyo, p. 268-269. http://csdl2.computer.org/comp/proceedings/icais/2002/1733/00/17330268.pdf.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.
3. Луценко Е.В. Автоматизированная система распознавания образов, математическая модель и опыт применения // В.И. Вернадский и современность (к 130-летию со дня рождения): Тезисы научно-практической конференции. – Краснодар: КНА, 1993. – С. 37–42.
4. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности: 351400 "Прикладная информатика (по отраслям)". – Краснодар: КубГАУ. 2004. – 633 с.
5. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие с грифами Министерства сельского хозяйства РФ и УМО для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп.– Краснодар: КубГАУ, 2006. – 615 с.
6. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). Монография (научное издание). – Краснодар: КЮИ МВД РФ, 1996. – 280с.
7. Луценко Е.В. АСК-анализ как метод выявления когнитивных функциональных зависимостей в многомерных зашумленных фрагментированных данных / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №03(11). – Режим доступа: http://ej.kubagro.ru/2005/03/pdf/19.pdf
8. Луценко Е.В. Виртуализация общества как основной информационный аспект глобализации / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №01(9). – Режим доступа: http://ej.kubagro.ru/2005/01/pdf/02.pdf
9. Луценко Е.В. Количественные меры возрастания эмерджентности в процессе эволюции систем (в рамках системной теории информации) / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2006. – №05(21). – Шифр Информрегистра: 0420600012\0089. – Режим доступа: http://ej.kubagro.ru/2006/05/pdf/31.pdf
10. Луценко Е.В. Критерии реальности и принцип эквивалентности виртуальной и "истинной" реальности / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. –

[1] Для обеспечения доступа читателей к этим и другим работам они размещены в Internet по адресам: http://lc.kubagro.ru/aidos/ http://ej.kubagro.ru/a/viewaut.asp?id=11

Краснодар: КубГАУ, 2004. – №06(8). – Режим доступа: http://ej.kubagro.ru/2004/06/pdf/10.pdf

11. Луценко Е.В. Математический метод СК-анализа в свете идей интервальной бутстрепной робастной статистики объектов нечисловой природы / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №01(3). – Режим доступа: http://ej.kubagro.ru/2004/01/pdf/13.pdf
12. Луценко Е.В. Программная идея системного обобщения математики и ее применение для создания системной теории информации / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №02(36). – Шифр Информрегистра: 0420800012\0016. – Режим доступа: http://ej.kubagro.ru/2008/02/pdf/11.pdf
13. Луценко Е.В. Неформальная постановка и обсуждение задач, возникающих при системном обобщении теории множеств на основе системной теории информации (Часть 1-я: задачи 1-3) / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №03(37). – Шифр Информрегистра: 0420800012\0031. – Режим доступа: http://ej.kubagro.ru/2008/03/pdf/12.pdf
14. Луценко Е.В. Неформальная постановка и обсуждение задач, возникающих при системном обобщении теории множеств на основе системной теории информации (Часть 2-я: задачи 4–9) / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №04(38). – Шифр Информрегистра: 0420800012\0049. – Режим доступа: http://ej.kubagro.ru/2008/04/pdf/03.pdf
15. Луценко Е.В. Семантическая информационная модель СК-анализа / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №02(36). – Шифр Информрегистра: 0420800012\0015. – Режим доступа: http://ej.kubagro.ru/2008/02/pdf/12.pdf
16. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(1). – Режим доступа: http://ej.kubagro.ru/2003/01/pdf/11.pdf
17. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(5). – Режим доступа: http://ej.kubagro.ru/2004/03/pdf/04.pdf
18. Луценко Е.В. Универсальный информационный вариационный принцип развития систем / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(41). – Режим доступа: http://ej.kubagro.ru/2008/07/pdf/10.pdf
19. Луценко Е.В. Численный расчет эластичности объектов информационной безопасности на основе системной теории информации / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(1). – Режим доступа: http://ej.kubagro.ru/2003/01/pdf/05.pdf
20. Пат. № 2003610986 РФ. Универсальная когнитивная аналитическая система "ЭЙДОС" / Е.В.Луценко (Россия); Заяв. № 2003610510 РФ. Опубл. от 22.04.2003. – 50с.
21. Пат. № 940217. РФ. Универсальная автоматизированная система распознавания образов "ЭЙДОС". /Е.В.Луценко (Россия); Заяв. № 940103. Опубл. 11.05.94. – 50с.
22. Перегудов Ф.И., Тарасенко Ф.П. Введение в системный анализ: Учебное пособие. – М.: Высшая школа, 1997. – 389с.