

Г. Н. Зверев, доктор технических наук
Уфимский государственный авиационный технический университет

**МОДЕЛИ НЕОПРЕДЕЛЕННОСТЕЙ
И ФУНДАМЕНТАЛЬНЫЕ КРИТЕРИИ ИНФОРМАТИКИ'**
[“Информационные технологии”, № 6, 2000.– С. 2–10]

Строится конструктивная семантика различных представлений формализованных неопределенностей – индефиниций, согласованных с фундаментальными критериями информационно-материальной деятельности: обобщенной сложности–простоты и ценности решений информационных задач либо риска потерь материальных, духовных, пространственно-временных ресурсов.

Кардинальные усовершенствования информационных языков и технологий во многом связываются в настоящее время с формализацией семантики основных понятий информатики, с уточнением и детализацией их “смысловой начинки”, определяющей в конечном итоге построение алгоритмов и программно-аппаратных средств. Одно из таких не вполне формализованных явлений информационного мира – это восприятие и понимание состояния неопределенности. Цель данной работы – построить формальную семантику различных моделей и мер неопределенных информационных ситуаций, существенно влияющих на языковую и алгоритмическую форму информационных процессов, найти связи между ними и увязать их с проблемой синтеза критериев материально-информационной деятельности.

Понятия неопределенности и информации являются исходными в информатике, они обычно считаются интуитивно ясными и очевидными, выражающими, соответственно, негативный и позитивный смысл информационных объектов и процессов, однако уровень формализации их семантики и распространенности в научно-технических текстах, весьма различен. Понятие информации, которую несет с собой некая знаковая структура – информационный объект x из пространства возможных значений X , детально описано в многочисленных публикациях. Выделим три основных свойства информационного объекта x :

– **объем** информации, измеряемый в произвольных единицах ее носителя – в битах или мегабайтах машинной памяти, в тоннах веса архивных документов, в минутах воспроизведения аудио- и видео- информации, в печатных листах издательской продукции и т.д.;

– **качество** информации, **информативность** объекта x , **количество полезной** информации, содержащейся в объекте x относительно целевого объекта y , измеряемое достигаемой точностью, погрешностью, остаточной неопределенностью цели y при использовании знания x ;

– **ценность**, материальная (экономическая и т.п.) польза или потери от использования знания либо незнания состояния x .

Неопределенность как научный термин – антипод понятию информации – имеет гораздо более скромное распространение и уровень формализации, хотя меры количества полезной информации, введенные Р. Хартли и К. Шенноном в теорию связи и передачи сообщений, в теорию информации [1], основаны на логарифмической мере неопределенности – информационной энтропии.

В общепринятом понимании слово “неопределенность” в описаниях знаковых ситуаций (в наблюдении, диагностике, прогнозировании, планировании, проектировании, управлении) означает присутствие в них неизвестных, искаженных, размытых объектов, незнакомых, запутанных явлений и неполных или противоречивых данных о них. Задача уточнения смысла и формализация семантики неопределенности как научного термина осложняется тем, что в нем объединяются, казалось бы, несовместимые свойства информационных категорий: знание и незнание, строгие, однозначные и обтекаемые, размытые представления объектов и процессов. Иными словами, термин “неопределенность” является неопределенным в обыденном значении этого слова, нам же крайне необходимо получить в результате формализации точно определенный информационный объект – термин с однозначно заданными типами неопределенностей и их моделями.

Соотношением между понятиями информации и неопределенности и их формализацией занимались многие ученые, см. [1-7]. Некоторые авторы приводят длинные списки терминов, выражающих различные аспекты представлений о неопределенности, подтверждая тем самым важность решения этой проблемы. В самом деле, формализация семантики этого понятия способствует уточнению смысла основных междисциплинарных терминов и общеупотребительных слов, таких как незнание, неизвестный или произвольный объект, неточность, недостоверность, ошибочность, противоречивость, неоднозначность, разнообразие, изменчивость, вариативность, хаотичность, случайность, нечеткость, невозможность, невероятность, а также противоположные им позитивные формы выражения свойств информационных объектов, такие как точность, однозначность, вероятность, шанс и т.д. Чаще всего неопределенность в научных текстах отождествляется с понятиями случайности и нечеткости [6].

Некоторые из этих терминов имеют одну или несколько различных формализаций в виде соответствующей меры и модели ее порождения или определения, привязанной к информационным средствам сравнения, измерения, вычисления, к фактическим и априорным источникам информации, к способам последующего использования моделей и мер неопределенности, другие пока остаются в виде субъективных конструкций, свободных от проверки мнений. Вполне формализованными являются понятия точности, погрешности, достоверности, вариативности, допустимости. Понятия вероятности и случайности, уточненные Мизесом и Колмогоровым в 20-х годах, через сорок лет приобрели законченную форму в алгоритмической теории информации Колмогорова и алгоритмической теории вероятностей Колмогорова и Мартин-Лёфа, см. обзор [8]. Понятие нечеткости (fuzzyness), введенное Л. Заде, строго говоря, не является конструктивным, так как определяющая его функция принадлежности объекта нечеткому множеству (fuzzy set) не увязана с источниками фактической и априорной информации, со

способами оценки последствий размытости объектов: ожидаемых потерь, точности, ценности решений.

Построение концептуальной или семиотической модели какого-либо явления обычно предшествует созданию математической модели. Следуя работе [9], будем отличать не вполне формализованные неопределенности от строго формализованных, получивших название **индефиниций**, которые при увеличении полезной информации и ее точности, при сжатии области неопределенности в одну точку превращаются в полную определенность – **дефиницию** или **идентификацию** объекта. Далее формализованные модели неопределенности – индефиниции строятся в семиотическом и системологическом базисе [10].

В семиотическом базисе каждый информационный объект – понятие Π , однозначное или размытое, представляется четырехслойной структурой: $\Pi=ИКАД$, в которой $И$ есть **имя**, обозначение объекта либо множество имен-синонимов: полное, краткое имя, идентификатор, первичный ключ и т.п.; $Д$ есть **дент** знака Π (его прямое значение, комодель, денотат, прообраз), выражающий объем понятия в виде константы или индивида $Д=const$, переменного объекта $Д=var$, либо множества значений – домена $Д=dom$; $К$ есть **конт** знака (его косвенное значение, модель, концепт, коннотат, образ объекта), выражающий содержание понятия в виде набора свойств K^0 и связей K^* объекта с другими объектами; $А$ – динамический информационно-материальный **адрес** знака Π , который содержит адреса всех его компонентов и обеспечивает пространственно-временные связи знаков информационного процесса. Имя, дент, конт и адрес определяются в информационной системе (точно, приближенно или остаются частично определенными) четырьмя дефинитивными функциями: именованная, дентированная, контированная и адресация.

В системологическом базисе информационный объект Π представлен структурным и ролевым описаниями, которые порождают соответствующие индефиниции. В ролевом системном аспекте объекты подразделяются на три вида:

- **функциональные** f -объекты, выполняющие в системе преобразования и их цепочки – информационные процессы;
- **статусные** s -объекты, описывающие состояния входных, внутренних и выходных объектов информационного процесса;
- **реляционные** r -объекты, определяющие связи между функциональными, статусными и самими реляционными объектами.

К ним добавляются объекты, выполняющие смешанные роли: ориентированные fs -объекты процедурной семантики, rs -объекты декларативной семантики и общий случай fsg -объектов процедурно-декларативной семантики математических моделей.

Семиотическое представление $ИКАД$ информационного объекта может быть постоянным или переменным, однозначно заданным или неопределенным. Поэтому необходимо различать **математические** и **семиотические** переменные. Математические переменные – функциональные, реляционные, статусные – имеют однозначно определенные имена, конты (типы переменных) и адреса, а дент принимает значения из вполне определенного домена, множества или пространства значений X –

функционального пространства $\{f\}$, пространства моделей связей $\{r\}$, числового или нечислового пространства состояний $\{s\}$.

Семиотические переменные характеризуются неопределенностями не только дента, но и других компонентов знака, его синтаксической $I=var$ и семантическими $КАД=var$ неопределенностями. Семиотические переменные играют важную роль в теории искусственного интеллекта, которая пока весьма далека от завершения. Здесь мы ограничимся изучением индефиниций математических переменных, точнее, элементарных статусных объектов – числовых (натуральных, целых, рациональных, вещественных) и нечисловых (номинативных, ранговых, иерархических, толерантных, циклических, логических) объектов состояния. Индефиниции составных статусных, функциональных, реляционных и семиотических переменных строятся по той же схеме, см. [10].

Итак, пусть некоторый материальный или информационный объект охарактеризован свойством x , числовым (длина объекта $x_1=1$ м, файл содержит $x_2=100$ записей) или нечисловым (цвет объекта $x_3=$ белый). Математическая переменная x по определению принимает значение из однозначно заданного множества возможных значений – пространства X . Формализованную неопределенность $ind(x)$ – индефиницию значения переменной x определим как понятие, обобщающее известные формализации междисциплинарных терминов неопределенности, перечисленных выше:

Индефиниция $ind(x)$ = неизвестность, незаданность +
+ вариативность, изменчивость +
+ искаженность, ошибочность +
+ многозначность, противоречивость,

где + означает объединение контов и дентов терминов при конкретизации и ограничении смыслов индефиниций в некоторой информационной ситуации. Раскроем содержание всех четырех составляющих данного определения.

Виды неопределенностей разделим, прежде всего, по числу m источников информации о свойстве x – по числу заданных значений x . Здесь под источником информации понимается измерительный прибор, алгоритм вычисления x по известным фактическим и априорным данным и т.п. Если $m=0$, то значение переменной не задано, оно не известно, но априори всегда нечто известно о возможных значениях x и их допустимых изменениях, скажем, известно подмножество Q_x допустимых значений $x \in Q_x \subset X$. Такая информационная ситуация характеризуется **вариативной индефиницией** $ind(x) = var(x)$, а пространство X пополняется знаком θ – крин, КРуглым Информационным Нулем, знаком внутренней неопределенности возможных значений, который включается в пространство возможных значений переменной x : $X_\theta = X + \{\theta\}$, $x \in X_\theta$, и при $x = \theta$ – “ x неизвестно” – знак θ отсылает информационный процесс к соответствующей модели неопределенности переменной x , к ее модели вариативности, см. далее.

Пусть имеется один источник информации $m=1$, т.е. задано одно значение переменной, скажем $x=1$ м или $x=$ белый, тогда информационная ситуация характеризуется адеквативной (точностной) индефиницией $ind(x)=adq(x)$, описывающей возможные отклонения известного значения выходного объекта \hat{x} источника информации от неизвестного истинного

значения x . Для числовых переменных неопределенность значения x характеризуется вариативностью $\text{var}(\Delta)$ погрешности источника данных $\Delta = \hat{x} - x$, в общем случае для нечисловых и числовых переменных – моделями условной вариативности $\text{var}(\hat{x}|x)$ либо $\text{var}(x|\hat{x})$, в которых значение \hat{x} , определяемое источником, известно, а истинное значение неизвестно, $x = \theta$.

В открытых системах и информационных ситуациях источник может присвоить переменной x несвойственное ей значение $\hat{x} = a \in X$ или $a \in Q$, противоречащее априорике. Такого вида информационные ситуации помечаются символом **внешней** неопределенности \square – квин, Квадратным Информационным Нулем. Знак квин при $x = \square$ указывает на наличие внешней фатальной ошибки, противоречия между источником информации и априорикой.

Другой вид противоречивой неопределенности возникает в информационных ситуациях с несколькими источниками значений переменной x , $m \geq 2$. Это случай согласованной или противоречивой многозначности значения x , обусловленной адеквативной вариативностью каждого источника, при этом противоречия могут возникнуть как между фактическими источниками, так и с априорикой. В первом случае $x = a, x = b, \dots, a \neq b$, во втором $a \in Q_x, \dots$, которые также помечаются знаком квин.

При формализации источников – сенфóров (сенсóров и рефóров) используются вариативные и адеквативные индефиниции. Здесь мы ограничимся описанием более простых индефиниций – моделей вариативности, т.к. модели адеквативности требуют описания не только реального источника – сенфора объекта \hat{x} , но и идеального сенфора, порождающего истинные значения x . Простейшей моделью вариативной неопределенности является множество $Q_x \subset X$ возможных, допустимых значений переменной $x \in Q_x$ – **аналитическая индефиниция** фон Неймана [4].

Если $Q_x = X$ – ситуация полной неопределенности, то индефиниция Q_x становится неинформативной. Другой крайний случай – область неопределенности Q_x стягивается в точку – синглетон, это ситуация полной определенности, однозначности, предельного “априорного зажатия”, при котором переменная x превращается в константу, и если предельная индефиниция Неймана известна точно, то эта ситуация есть безошибочная **идентификация** переменной, полная определенность и адекватность значения x . Все другие ситуации характеризуются вариативностью и неадекватностью значения переменной x . Если же множество Q_x пустое, то индефиниция и выраженная в ней априорика противоречива.

Второй тип моделей вариативности составляют **сомножества** (комплекты, мультимножества, фактормножества). В отличие от множества в сомножестве допускаются неразличимые элементы. Сомножество Q_x^c задается в виде совокупности несвязанных, частично различимых элементов из X , полученных путем многократного выбора, копирования, размножения,

объединения элементов в классы эквивалентности и абстрагирования от именованности и адресации элементов либо выбора единственного “типичного” представителя каждого класса эквивалентности. Сомножество определяется заданием на X числовой функции распределения численности $N(x)$ или частоты $q(x) = N(x)/N$ неразличимых элементов сомножества, $N = |Q_x^c|$ – мощность сомножества, равная $\sum_x N(x)$.

Понятие сомножества и его функции распределения более соответствует описанию размытых информационных ситуаций, чем понятие множества, т.к. сомножество однозначно определяет не только принадлежность, но и то, как часто могут появляться те или иные значения переменной $x \in X$ на сомножестве Q_x^c и если $N(x) = 0$, то данное значение не принадлежит сомножеству и является недопустимым в данной априорике, $x = \square$. При абстрагировании от численности и частоты, оставляя лишь сам факт принадлежности элемента сомножеству: $N(x), q(x) > 0$, и сливая неразличимости в один элемент, получают из сомножества Q_x^c множество Q_x , из функции численности и частоты – двоичную функцию принадлежности аналитической модели Неймана.

Вариации переменной x могут быть обусловлены природными или информационными причинами, свойствами объекта, множества объектов или источника информации. Изменения значения x , соответствующие одному или множеству объектов, могут быть детерминированными или стохастическими, случайными. В последнем случае сомножество Q_x^c и распределение частоты $q(x)$ приобретает теоретико-вероятностную семантику: сомножество есть генеральная совокупность, а распределение частоты есть распределение вероятности появления случайного значения x . При случайных независимых вариациях Q_x^c и $q(x)$ превращаются в модели полной априорной информации, для детерминированных, закономерных изменений это не так, см. далее.

Третий тип вариативностей выражают **траекторные** (исторические) модели T_x переменной x в виде последовательности ее значений $(x_i)_N = x_1, x_2, \dots, x_N$, т.е. траекторная модель есть линейно упорядоченное сомножество, несущее дополнительную информацию об истории вариаций переменной x , о динамике сенфорного процесса.

Четвертый тип моделей вариативностей составляют алгоритмические модели генерации последовательностей – геноры Γ_x , определяющие механизмы порождения значений переменной x . **Генор** Γ_x есть функциональный объект, замкнутый по входу, порождающий последовательно значения x , он представляет собой **модель полной информации** о переменной x и служит основным объектом метода Монте-Карло, а также алгоритмических теорий информации и вероятностей [8]. Множество и сомножество характеризуют статическую составляющую завершённой или ожидаемой вариативности, траектория и генор дополнительно описывают динамику вариаций переменной x .

Аналитическая Q_x , частотная $q(x)$, траекторная T_x , алгоритмическая Γ_x модели вариативной неопределенности образуют иерархию априорной информации о переменной x . Вершиной иерархии служит алгоритмическая

модель полной информации Γ_x , которая описывает все детали механизма вариативности, историю изменения, частоту появления и множество возможных значений переменной x . Отвлекаясь от механизма порождения или очередного изменения значения x , мы теряем алгоритмическую составляющую индефиниции $\text{ind}(x) = \Gamma_x$ и сохраняем лишь результат функционирования алгоритма Γ_x – последовательность T_x . Абстракция от порядка изменения значений x приводит к потере информации о характере вариаций, резких или плавных, детерминированных или случайных, остается лишь результат варьирования – множество неупорядоченных значений x и его распределение частоты $q(x)$ для детерминированных и случайных вариаций либо вероятности случайных значений в теоретико-вероятностном смысле. Абстракция от численности (частоты, вероятности) переводит множество в множество отбрасыванием всех повторов элементов, при этом частотные связи теряются либо превращаются в логические, числовая функция распределения $q(x)$ – в двоичную функцию принадлежности, несущую минимальную информацию о допустимости значений переменной x .

Перевернув информационную иерархию априорика $Q_x \rightarrow q(x) \rightarrow T_x \rightarrow \Gamma_x$ и изменив порядок на обратный, получим иерархию индефиниций $\Gamma_x \rightarrow T_x \rightarrow q(x) \rightarrow Q_x$, в основании которой лежит полная априорика переменных, а на вершине – наиболее бедная модель неопределенности – аналитическая индефиниция классической математики. Если вариации переменной x в области неопределенности Q_x являются случайными в смысле Колмогорова–Мизеса с неизвестным механизмом Γ_x порождения очередного значения переменной, имеющей максимально сложную по Колмогорову траекторию, то извлечь информацию из хаоса T_x по определению случайности [8] не представляется возможным, поэтому моделью полной информации стохастических процессов, как это принято в математической статистике, является распределение вероятностей $q(x)$, так как попытки извлечь полезную информацию из частичного знания случайного механизма Γ_x или из траекторной модели T_x лишь ухудшают информационные характеристики процесса решения, а полное знание алгоритмической модели Γ_x вовсе устраняет случайность и неопределенность, превращая их в детерминированную предопределенную однозначность.

Индефиниции $Q_x, q(x), T_x, \Gamma_x$ описывают неопределенность индивидуального объекта x однозначным заданием порождаемого моделью класса возможных значений переменной x и свойств этого класса в виде различных мер неопределенности, а через них, в конечном итоге, – критерии эффективности информационных процессов. Источниками неопределенности и ее моделей могут быть: 1) изменчивость реального объекта и окружающей его среды или множества объектов – универсума информационных ситуаций предметики, для которого строят информационные процессы исследования, проектирования и управления; 2) неконтролируемая изменчивость информационной среды – ее средств измерения и знаковых преобразований; 3) размытость и неадекватность исходной формализации, фактической и априорной информации.

Пусть нам задан универсум объектов предметики $U = \{obj\}$ или множество соответствующих им информационных ситуаций мощностью $N = |U|$, а также средство выбора объектов из универсума – генор Γ_0 и сенфор AB , содержащий средство наблюдения – сенсор A и средство обработки информации – рефор B , на выходе которого образуется значение переменной $x: U \xrightarrow{\Gamma_0} obj \xrightarrow{AB} x$, тогда вариативность x есть следствие разнообразия объектов универсума U по свойству x , разнообразия выбора объектов моделью Γ_{obj} , точности и чувствительности сенфора AB . Меры разнообразия и неопределенности значения переменной x зависят также от типа шкалы X рефора B .

Для дискретной переменной $x \in X$, числовой или нечисловой, значность шкалы $k = |X|$ определяет потенциальное разнообразие значений x , а мощность аналитической модели $k_0 = |Q_x|$ определяет **фактическую значность** переменной x , порождаемой ее генором $\Gamma_x = U \cdot \Gamma_0 \cdot AB$ – это алгоритмическая модель неопределенности x в аппликативной записи.

Счетные меры неопределенности аналитической индефиниции Q_x можно определить в различных шкалах мер. Естественные меры с ясной семантикой определяются:

- в линейной шкале счета возможных альтернатив разнообразия значений – это **предельный альтернант** $L_x^0 = k_0 - 1$ переменной x ;
- в логарифмической шкале счета числа альтернатив – это **энтропия** Хартли $H_x^0 = \log k_0$ переменной x , равная предельному уровню меры Шеннона.

Альтернант и энтропия равны нулю при $k_0 = 1$ – случай полной определенности значения $x \in Q_x$ – одноточечному множеству, а в общем случае мера разнообразия k_0 изменяется в пределах $1 \leq k_0 \leq k$, $Q_x \neq \emptyset$, предельные счетные меры аналитической индефиниции удовлетворяет ограничениям: $0 \leq L_x^0 \leq k - 1$, $0 \leq H_x^0 \leq \log k$

Счетные меры сомножества Q_x^c и его распределения $q(x)$, траектории T_x и генора Γ_x определяются как средние значения противоположной или обратной редкости значения переменной в линейной или логарифмической шкале счета [10]: 1) **альтернант** $L_x = k_0 \sum_{i=1}^k q_i \cdot \bar{q}_i = k_0 \sum_{i=1}^k \sigma_i^2$, где q_i – частость значений x_i , $\bar{q}_i = 1 - q_i$ – его противоположная редкость, $\sigma_i^2 = q_i(1 - q_i)$ – дисперсия двоичной переменной $x_i \in Bit$; 2) **модифицированный альтернант** $L'_x = (\sum_{i=1}^k \sigma_i)^2$; 3) **энтропия Шеннона** $H_x = -\sum_{i=1}^k q_i \log q_i = \sum_{i=1}^k q_i \log q_i^{-1}$ – среднее значение логарифмической редкости – логарифма обратной редкости $1/q_i$.

Альтернанты (1,2) и энтропия (3) изменяются в пределах $0 \leq L_x$, $L'_x \leq L_x^0$, $0 \leq H_x \leq H_x^0$ и верхняя граница достигается при равномерном распределении $q(x)$, порожденном моделями T_x, Γ_x , когда все альтернативы

становятся равноправными по частоте или равновероятными – это случай максимальной неопределенности. Единица измерения альтернанта – альт, а величины L_x , L'_y показывают в среднем, сколько необходимо отбросить альтернатив, чтобы достичь полной определенности. Единицы измерения энтропии определяются основанием логарифма – бит, бел, децибел, а величина H_x показывает в среднем длину кода значений x в соответствующем алфавите, который обеспечивает однозначную различимость ожидаемых значений переменной x .

Неопределенность вещественной переменной бесконечна, поэтому вводятся ограничения диапазона вариаций $a \leq x \leq b$, исходя из уровня доверия – риска, и порог различимости $\Delta_{\Pi}(x)$, в общем случае зависящий от x . Альтернант плотности распределения $q(x)$, траектории T_x , генерации Γ_x вещественных значений x с ограниченной (порогом) разрядностью представления x есть $L_x = k_0(1 - \int q^2(x) \cdot \Delta_{\Pi}(x) dx)$, $k_0 = \int_a^b \frac{dx}{\Delta_{\Pi}(x)}$,

модифицированный альтернант $L'_x = (\int \sqrt{q(x)\Delta_{\Pi}^{-1}(x) - q^2(x)} dx)^2$, энтропия $H_x = -\int q(x) \log(q(x)\Delta_{\Pi}(x)) dx$. Они имеют в конечном мире ту же семантику, что и для дискретной переменной.

Для числовых переменных, дискретных и непрерывных, помимо счетных мер неопределенности вводятся метрические меры: радиус R_x , диаметр D_x , среднеквадратический размер σ_x области неопределенности Q_x^c и другие меры, причем для вычисления D_x достаточно минимальной информации – Q_x , а для определения R_x и σ_x необходимо знать распределение $q(x)$: $R_x = \max_x |x_0 - x|$, $D_x = \max_{x_1, x_2} |x_1 - x_2|$, $\sigma_x^2 = \int (x - x_0)^2 q(x) dx$, $x_0 = \int xq(x) dx$ – среднее значение, центр области рассеяния Q_x , который иногда определяют как моду, медиану и иным способом.

Между мерами неопределенности L_x , H_x , R_x , D_x , σ_x существуют тесные связи, которые позволяют при неизменной различимости Δ_{Π} либо заданной зависимости $\Delta_{\Pi}(x)$ предсказывать по какой-либо известной мере остальные меры неопределенности. Параметры связи между ними называются

факторами распределения: пиковые факторы. $t_q = \frac{R_x}{\sigma_x}$ – **радиальный**

пикфактор (коэффициент Стьюдента), который показывает во сколько раз пиковое значение вариации x – максимальное отклонение от центра рассеяния больше среднего отклонения σ_x при заданном доверии $q_c \leq 1$ или риске

$q_r = 1 - q_c \geq 0$, $t_q \geq 1$ и $\tau_q = \frac{D_x}{\sigma_x}$ – **диаметральный пикфактор**, $\tau_q > 1$;

$\delta_q = \frac{D_x}{\Delta_{\Pi}} - L_x$ – **сдвиговый фактор**, $\delta_q \geq 1$; $\varepsilon_q = \frac{\Delta_{\Pi}}{\sigma_x} 2^{H_x}$ – **энтропийный**

фактор, энтропия измеряется в битах, $\varepsilon_q \leq \sqrt{2\pi e} \approx 4.13$; $\gamma_q = 2 - \frac{D_x}{R_x}$ – **фактор**

асимметрии распределения, $0 \leq \gamma_q < 1$. Зная факторы распределения можно вычислить разные меры неопределенности по формулам : $R_x = t_q \cdot \sigma_x$,

$D_x = \tau_q \cdot \sigma_x$, $L_x = \frac{D_x}{\Delta_{II}} - \delta_q$, $H_x = \log\left(\frac{\sigma_x}{\Delta_{II}} \varepsilon_q\right)$, $D_x = (2 - \gamma_q)R_x$. Факторы

распределений слабо зависят от параметров ожидаемого универсума объектов предметики и обычно имеют достаточно узкий диапазон вариаций.

Для динамических индефиниций T_x , Γ_x помимо введенных выше мер неопределенности вводятся характеристики регулярности и иррегулярности вариаций, частотные и метрические свойства траекторий и алгоритмов генерации значений числовых и нечисловых элементарных переменных. Для составных статусных объектов – векторов, матриц, функциональных и реляционных объектов модели и меры неопределенности усложняются, но их семантика, в основном, повторяет изложенные выше формализации, см. [10]. В этой работе приводятся формулы преобразований моделей и мер неопределенностей в дискретных и непрерывных информационных процессах.

Как и всякая информация, модели и меры неопределенности могут быть заданы не полностью, а лишь частично, информация при этом может искажаться: $\hat{Q}_x = Q_x + \Delta Q_x$, $\hat{q}(x) = q(x) + \Delta q(x)$ и т.п., подвергаться вариациям и размытиям в силу внешних причин, но это уже неопределенности следующего иерархического уровня. Иерархические модели неопределенности «свертываются» в актуальную действующую индефиницию замкнутой системы, например, вариации распределения $q(x)$, обязанные неопределенностями его параметров z определяют «свертку» $q_s(x) = \int q(x|z)q(z)dz$ как итоговую неопределенность [11].

Точную частичную информацию о модели неопределенности означим кавычками, скажем, « Q_x »= (x_0, σ_x) , Q_x =« Q_x^c », T_x =« Γ_x », $q(x)$ =« T_x » и т.п., а полную или частичную искаженную информацию будем помечать сверху шляпкой: $\hat{\Gamma}_x$ и т.д., например, сомножество частично описывается двумя моментами распределения « $q(x)$ »= (x_0, σ_x^2) – центром рассеяния и ожидаемыми квадратическими размерами, а того же типа искаженная частичная априорика о сомножестве есть « $\hat{q}(x)$ »= $(\hat{x}_0, \hat{\sigma}_x^2)$.

Тем самым в информационную семантику неопределенностей помимо иерархии вариативностей $\Gamma_x \rightarrow T_x \rightarrow q(x) \rightarrow Q_x$ вводятся: иерархия полноты описания их свойств $(x_0, \sigma_x, t_q, \dots)$, иерархия их адекватностей или погрешностей их задания $(\hat{x}_0, \hat{\sigma}, \dots)$. Не вполне заданные и искаженные индефиниции переменной x или ее погрешности $\Delta = \hat{x} - x$ часто заменяются точными или приближенными оценками сверху и снизу.

В общем случае стандартными формами задания ограничений вариативности числовых моделей ($\hat{x}_1 = s \pm \Delta s$, $\hat{x}_2 = f \pm \Delta f$, $\hat{x}_3 = r \pm \Delta r$) служат r -объекты различных типов: 1)уравнения, 2)неравенства, 3)экстремальные критерии оптимизации, 4)частотные или вероятностные распределения, а также

комбинации (1-4). Первые три определяют строгие, последние – размытые ограничения, в пределе описывающие строгие критериальные условия информационных задач. Последние также часто бывают размытыми и подлежат уточнениям в метазадачах синтеза критериев материально-информационных процессов.

Информационная деятельность есть преодоление неопределенностей самого разного типа при оценках текущего состояния, восстановлении прошлого и предсказании будущего в условиях ограниченных материальных и информационных ресурсов. Размытость и ошибочность фактических и априорных данных ухудшает качество информационных средств и технологий, которое оценивается набором показателей и критериев. Развитие теории сложности, теории статистических решений и теории полезности позволили продвинуться в решении проблемы синтеза критериев, обобщить теоретические подходы и практические приемы [9,11], сформулировать фундаментальные критерии информатики.

Разделение объектов и процессов на материальные и информационные (знаковые) упрощает задачу синтеза путем декомпозиции критериев на информационные и ценностные. В соответствии с теорией обобщенной сложности – простоты информационные критерии определяют достижимость и меры удаленности от поставленных целей при заданных информационных ресурсах, а ценностные критерии дополнительно учитывают более общие жизненные аспекты, ограничения всех значимых ресурсов, ожидаемые материальные, социальные и иные последствия принятых или отложенных решений при исследовании, проектировании, управлении. В теории сложности различают сводимые друг к другу статические меры сложности объектов и динамические меры сложности процессов. Далее в настоящей работе используются последние меры, а сложность объекта или системы взаимосвязанных объектов характеризуется сложностью алгоритма и процесса их анализа и синтеза. [10 - 12].

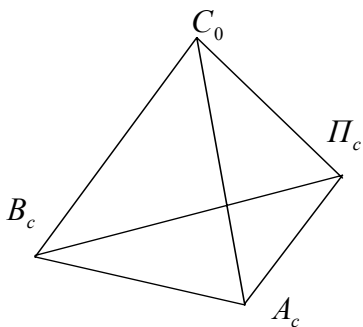
Меры сложности и простоты моделей объектов и процессов позволяют количественно характеризовать с достаточной полнотой и всесторонностью существенные свойства информационных задач: разрешимость, достоверность, своевременность, выраженные в обобщенной позитивной форме как критерии простоты либо в негативной форме как критерии сложности достижения поставленных целей. Компоненты обобщенной сложности–простоты являются фундаментальными критериями оценки информационной деятельности, выживаемости ее объектов и технологий.

Основными параметрами информационных критериев служат : **время B** или скорость B^{-1} – оперативность решения, получаемого информационной системой либо специалистом – предметником; **адекватность A** – точность, истинность либо погрешность, ошибочность A^{-1} результата; **память Π** , требуемый объем информации, информационная емкость либо компактность Π^{-1} знакового объекта, алгоритма, технологии. Наилучшие значения этих параметров, определяющих идеальную нереализуемую информационную систему или технологию, есть значения, ближайšie к нулевым: $B, \Pi, A^{-1} \cong 0$ – решение получается мгновенно, оно предельно компактно и достоверно, т.е. объем информации минимален, а количество полезной информации, информативность решения максимальны.

Если к этим параметрам добавить информацию о достигнутых в прототипах уровнях точности, компактности, оперативности решений, которая определяет границы B_{Γ} , A_{Γ} , Π_{Γ} между простотой и сложностью, то получим их относительные меры – фундаментальные информационные критерии:

- **временная** или операционная сложность $B_c = B/B_{\Gamma}$ и простота $B_{\Pi} = B_{\Gamma}/B$;
- **адеквативная** или точностная сложность $A_c = A_{\Gamma}/A$ и простота $A_{\Pi} = A/A_{\Gamma}$, которые в некоторых постановках информационных проблем, в частности, при вероятностно-статистическом подходе, заменяются относительными мерами вариативности в виде косвенных критериев точности, сложности и простоты;
- **объемная** или емкостная сложность $\Pi_c = \Pi/\Pi_{\Gamma}$ и простота $\Pi_{\Pi} = \Pi_{\Gamma}/\Pi$ процесса решения.

Обобщенная или интегральная сложность C_o является функцией от составляющих ее компонентов B_c , A_c , Π_c , скажем, в виде линейной взвешенной суммы – средней сложности: $C_o = p_B B_c + p_A A_c + p_{\Pi} \Pi_c$, $p_B + p_A + p_{\Pi} = 1$. Простым системам соответствуют единичные интервалы мер сложности : $0 < C_o < 1$, $0 < B_c < 1$ и т.д., а сложным объектам и процессам соответствуют меры, превышающие заданные граничные значения: $C_o, B_c, A_c, \Pi_c > 1$. Компоненты обобщенной сложности в значительной степени взаимозависимы и изменение одной из них, скажем, требуемой точности, приводит к изменению времени и информационной емкости процесса решения. Связи между фундаментальными критериями информатики изображаются в виде критериальной пирамиды – информационного СВАП-тетраэдра, в основании которого лежат базисные критерии сложности – простоты, а в вершине – совокупная сложность C_o , вычисляемая в общем случае по заданной функции сложности F_c и базисным характеристикам: $C_o = F_c(B_c, A_c, \Pi_c)$.

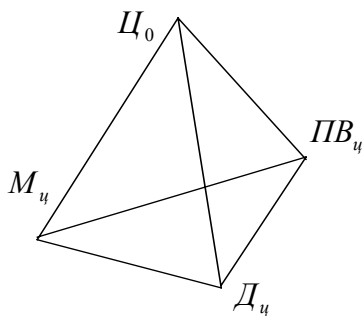


Введенные меры сложности применимы к информационным ситуациям с достижимой целью, с известным алгоритмом и процессом решения, в противном случае возникают неопределенности метрической алгоритмики и в информационный процесс включают метаалгоритмы формализации, построения алгоритмов решения, анализа их применимости и достижимости цели, что увеличивает общую сложность, время, память, либо вводят дополнительные критерии достижимости – меры **универсальности** (частичности); **надежности** информационного процесса, нарушение которых ведет к недостижимости типа крин (решение $\hat{x} = \theta$ – алгоритм не сходится, произошло заикливание,

исчерпаны ресурсы и т.п.) либо типа квин ($\hat{x} = \square$ – произошел сбой, есть противоречия в фактах, в априорике или алгоритме).

Более реалистическое описание информационных ситуаций и их ресурсов состоит в переходе к ценностным критериям, которые учитывают более широкий спектр целей и ограничений материального и информационного производства. Решение этой проблемы состоит в построении:

- шкал обобщенных ценностей, выражающих в позитивной форме меру полезности либо в негативной форме меру вреда – ожидаемых затрат, потерь, рисков;
- иерархии ценностей для каждого этапа материально-информационной деятельности в виде **аксиологических пирамид**.



Простым, но достаточно общим примером такой пирамиды служит ценностный ЦМДПВ-тетраэдр, подобный информационному СВАП-тетраэдру, в основании которого лежат базисные меры ценности, а вершиной служит обобщенная ценность C_0 , мера полезности человеческой деятельности, вычисляемая по **функции ценности** F_u и основным ценностным характеристикам $C_0 = F_u(M_u, D_u, PB_u)$. В этой формуле параметр M_u

есть мера материальной ценности в единицах экономических категорий прибыли, дохода, затрат либо материальных показателей экономии материалов, энергии и других ресурсов; D_u - мера духовной полезности решения, научной, культурной, эстетической, эмоциональной ценности результата, PB_u - пространственно-временная ценность, выраженная в экономии или расширении жизненного, физического пространства, размеров, площадей, объемов, экономии реального времени, уменьшении простоев техники, людей и т.п. Состав показателей, форма выражения, единицы измерения пользы и вреда, виды связей критериев в аксиологической пирамиде уточняются в каждой предметной области и информационной ситуации при синтезе критериев и ограничений [12].

Фундаментальные информационные критерии естественным образом преобразуются в ценностные показатели, для чего используются **модели сопряжения** F_{cu} информационных, материальных, социальных и т.д. технологий. Эти модели изображаются в виде покомпонентных преобразований скорости, точности и т.д. в стоимость, меру риска, абстрактную полезность, либо в общем виде: $F_{cu}: СВАП \rightarrow ЦМДПВ$. Тем самым вариативные и адеквативные индефиниции, их польза и вред определяются в шкалах ценности. Примерами моделей сопряжения являются формулы расчета байесовских рисков завышения или занижения оценок, ожидаемых платежей и платежных матриц в теории статистических решений.

Сами по себе адеквативность, оперативность, компактность, универсальность, надёжность не являются основными целями информационной деятельности. Так, точность может быть излишней, да и сами знания, достоверные и не очень, могут быть ценными, бесполезными либо вредными, что учитывается опытным специалистом при синтезе решения. При этом важно

иметь объективные средства оценки последствий вариативных или адеквативных неопределённостей в шкалах ценности или простоты и знать, насколько априорная информация в виде индефиниций улучшает показатели информационных технологий и как субъективные или ошибочные характеристики фактов и моделей могут ухудшать оценки либо делать их недостоверными.

Подведем итоги. Процесс решения информационной задачи направлен на уменьшение вариативной или адеквативной неопределённости целевых неизвестных. Алгоритм решения должен строиться с учётом базисных информационных критериев обобщённой сложности-простоты процесса решения и обобщенной ценности, которые учитывают точность, время, необходимую память, универсальность – меру области применимости алгоритма в виде ожидаемой частоты его применения, меру надёжности в виде частоты достижения алгоритмом целевого результата. Сложностные критерии зависимы между собой и определяют противоречивые требования, поэтому их сводят к более общим ценностным критериям.

Формализованная семантика неопределённости определяется типом переменной, моделью и мерами неопределённости. Выделяют следующие типы индефиниций:

- семиотические (семантические, синтаксические) и математические;
- функциональные, реляционные и статусные;
- вариативные и адеквативные;
- согласованные и противоречивые;
- детерминированные и вероятностные;
- природные и информационные;
- статические Q_x , $q(x)$ и динамические T_x, Γ_x , индефиниции и их вариативные и адеквативные иерархии. Каждая индефиниция определяет несколько мер неопределённости: альтернант L_x , энтропию H_x , радиус R_x , диаметр R_x , среднеквадратический размер σ_x области неопределённости.

Полезность изложенного к формализации неопределённости и эффективность применения индефиниций доказаны решением размытых вычислительных и дискретных задач, в частности, в работе [11] приводится оптимальное по точности линейное решение системы уравнений с размытой прямоугольной матрицей произвольного ранга, с конечной правой частью и размытой целью, более адекватно представляющее реальные информационные ситуации. Н. Винер считал информацию о возможном более ценной, чем информацию о действительном. Очевидно, индефиниции обеспечивают строгое описание мира возможностей и неопределённостей, к которым применимы точные математические методы и информационные технологии.

Список литературы

1. Теория информации и ее приложения.– М.: Гос. изд. физ.-матем.лит., 1959, 328 с.
2. Фишер Р.А. [Fisher R.A.] Theory of statistical estimation – Proc. Camb. Phil.Soc., v.22, 1925, pp. 700-725.
3. Сцилард Л. [Szilard L.] Zs/ Physik, 53, 1929, 840.

4. Нейман Дж.фон, Моргенштерн О. Теория игр и экономическое поведение. .– М.: Наука, 1970, 707 с.
5. Колмогоров А.Н. Три подхода к определению понятия «количество информации»././ Проблемы передачи информации, № 1, 1965, С 3-11.
6. Заде Л.А. (L.A. Zadeh) Outline of a New Approach to the Analysis of Complex Systems and Decision Processes – IEEE Trans. Syst., Man, Cybern., vol. SMC–3, 1973, Jan., pp. 28-44.
7. Нариньяни А.С. Недоопределенность в системе представления и обработки знаний. – Изв. АН СССР. Техн. кибернетика, № 5, 1986.С. 3-28.
8. Звонкин А.К., Левин Л.А. Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов. – УМН, т. 25, вып. 6., 1970, С. 85-127.
9. Зверев Г.Н. К обобщенной теории обработки наблюдений.// Нефтепромысловая геофизика. – М.:ИГ и РГИ, 1974, С. 3-51.
10. Зверев Г.Н. Основания теоретической информатики. Разд. 1–7 – Уфа, УГАТУ, 1995– 97.
11. Зверев Г.Н., Дембицкий С.И. Оценка эффективности геофизических исследований скважин. – М.: Недра, 1982, 224 с.
12. Зверев Г.Н. Семантические схемы информационных технологий.//Межвуз. сб. «Измерительные преобразователи и информационные технологии»– Уфа, изд. «Гилем» АН РБ, 1996, С. 148-158.