

Лекция 2: Основы анализа информации, необходимой для решения профессиональных задач

План:

1. АДЕКВАТНОСТЬ ИНФОРМАЦИИ И ЕЁ ФОРМЫ
2. СИНТАКСИЧЕСКАЯ МЕРА ИНФОРМАЦИИ
3. СЕМАНТИЧЕСКАЯ МЕРА ИНФОРМАЦИИ
4. ПРАГМАТИЧЕСКАЯ МЕРА ИНФОРМАЦИИ
5. КАЧЕСТВО ИНФОРМАЦИИ

1. Адекватность информации и её формы

Информация - сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределенности, неполноты знаний.

Для потребителя информации очень важной характеристикой является ее адекватность.

Адекватность информации - это определенный уровень соответствия создаваемого с помощью полученной информации образа реальному объекту, процессу, явлению и т.п.

В реальной жизни вряд ли возможна ситуация, когда вы можете рассчитывать на полную адекватность информации. Всегда присутствует некоторая степень неопределенности.

От степени адекватности информации реальному состоянию объекта или процесса зависит правильность принятия решений.

Формы адекватности информации

Адекватность информации может выражаться в трех формах: семантической, синтаксической, прагматической.

Синтаксическая адекватность.

Она отображает формально-структурные характеристики информации и не затрагивает ее смыслового содержания.

На синтаксическом уровне учитываются тип носителя и способ представления информации, скорость передачи и обработки, размеры кодов представления информации, надежность и точность преобразования этих кодов и т. п.

Информацию, рассматриваемую только с синтаксических позиций, обычно называют данными, т.к. при этом не имеет значения смысловая сторона.

Семантическая (смысловая) адекватность.

Эта форма определяет степень соответствия образа объекта и самого объекта. Семантический аспект предполагает учет смыслового содержания информации.

На этом уровне анализируются те сведения, которые отражает информация, рассматриваются смысловые связи.

В информатике устанавливаются смысловые связи между кодами представления информации. Эта форма служит для формирования понятий и представлений, выявления смысла, содержания информации и ее обобщения.

Прагматическая (потребительская) адекватность.

Она отражает отношение информации и ее потребителя, соответствие информации цели управления, которая на ее основе реализуется.

Проявляются прагматические свойства информации только при наличии единства информации (объекта), пользователя и цели управления.

Прагматический аспект рассмотрения связан с ценностью, полезностью использования информации при выработке потребителем решения для достижения своей цели.

С этой точки зрения анализируются потребительские свойства информации.

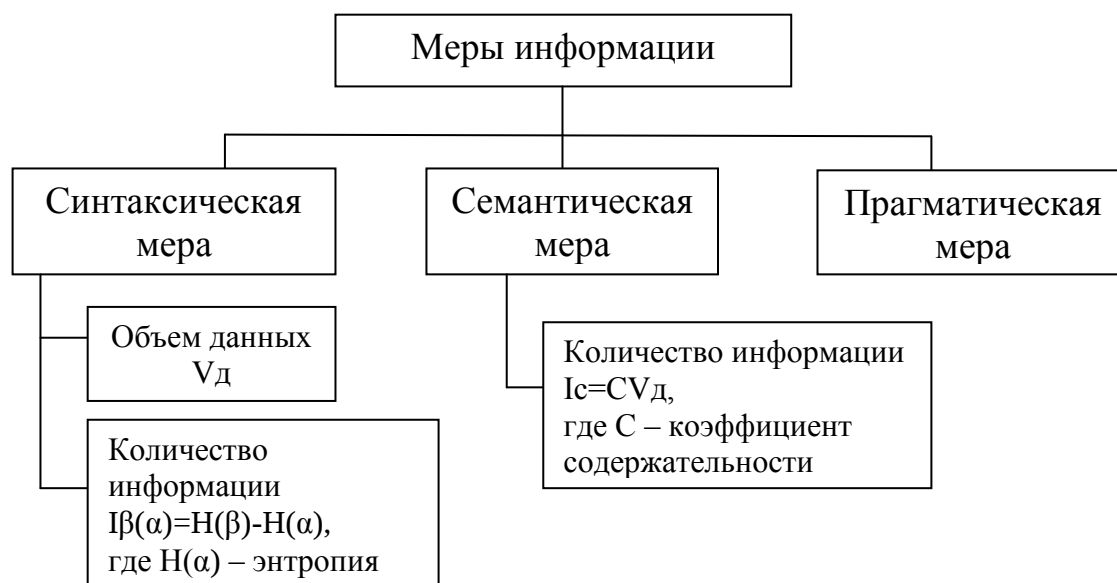
Эта форма адекватности непосредственно связана с практическим использованием информации, с соответствием ее целевой функции деятельности системы.

2. Синтаксическая мера информации

Для измерения информации вводятся два параметра: количество информации I и объем данных V .

Эти параметры имеют разные выражения и интерпретацию в зависимости от рассматриваемой формы адекватности.

Каждой форме адекватности соответствует своя мера количества информации и объема данных (рис):



Синтаксическая мера информации

Эта мера количества информации оперирует с обезличенной информацией, не выражающей смыслового отношения к объекту.

Объем данных V_d в сообщении измеряется количеством символов (разрядов) в этом сообщении.

В различных системах счисления один разряд имеет различный вес и соответственно меняется единица измерения данных:

- в двоичной системе счисления единица измерения - бит (bit - binary digit – двоичный разряд);
- в десятичной системе счисления единица измерения дит (десятичный разряд).

Количество информации I на синтаксическом уровне невозможно определить без рассмотрения понятия неопределенности состояния системы (энтропии системы).

Этот способ измерения количества информации впервые предложил в 1948 г. **К. Шеннон**.

По К. Шеннону, **информация** — это сведения, уменьшающие неопределенность (энтропию), существовавшую до их получения.

Действительно, получение информации о какой-либо системе всегда связано с изменением степени неосведомленности получателя о состоянии этой системы.

Рассмотрим это понятие.

Пусть до получения информации потребитель имеет некоторые **предварительные (априорные) сведения о системе α** .

Мерой его неосведомленности о системе является функция **$H(\alpha)$** , которая в то же время служит и мерой неопределенности состояния системы.

После получения некоторого **сообщения β** получатель приобрел некоторую **дополнительную информацию $I\beta(\alpha)$** , уменьшившую его априорную неосведомленность так, что **апостериорная (после получения сообщения β) неопределенность состояния системы стала $H\beta(\alpha)$** .

Тогда **количество информации $I\beta(\alpha)$** о системе, полученной в сообщении β , определится как **$I\beta(\alpha)=H(\alpha)-H\beta(\alpha)$** ,

т.е. количество информации измеряется изменением (уменьшением) неопределенности состояния системы.

Если **конечная неопределенность системы $H\beta(\alpha)$** обратится в нуль, то первоначальное неполное знание заменится полным знанием и количество информации **$I\beta(\alpha)=H(\alpha)$** .

Иными словами, **энтропия системы $H(\alpha)$** может рассматриваться как мера недостающей информации.

Энтропия системы $H(\alpha)$, имеющая **N** возможных состояний, согласно *формуле Шеннона*, равна:

$$H(\alpha) = - \sum_{i=1}^N P_i \log P_i,$$

где P_i - вероятность того, что система находится в i -ом состоянии.

Для случая, когда все состояния системы равновероятны, т.е. их вероятности равны

$$P_i = \frac{1}{N}$$

, ее энтропия определяется соотношением:

$$H(\alpha) = - \sum_{i=1}^N \frac{1}{N} \log \frac{1}{N}.$$

Часто информация кодируется числовыми кодами в той или иной системе счисления, особенно это актуально при представлении информации в компьютере.

Естественно, что одно и то же количество разрядов в разных системах счисления может передать разное число состояний отображаемого объекта, что можно представить в виде соотношения

$$N = m^n,$$

где N - число всевозможных отображаемых состояний;

m - основание системы счисления (разнообразие символов, применяемых в алфавите);

n - число разрядов (символов) в сообщении.

Наиболее часто используются двоичные и десятичные логарифмы. Единицами измерения в этих случаях будут соответственно бит и дит.

Коэффициент (степень) информативности (лаконичность) сообщения определяется отношением количества информации к объему данных, т.е.

$$Y = 1/V_d, \text{ причем } 0 < Y < 1$$

С увеличением Y уменьшаются объемы работы по преобразованию информации (данных в системе).

Поэтому стремятся к повышению информативности, для чего разрабатываются специальные методы оптимального кодирования информации.

3. Семантическая мера информации

Для измерения смыслового содержания информации, т.е. ее количества на семантическом уровне, наибольшее признание получила **тезаурусная мера**, которая связывает семантические свойства информации со способностью пользователя принимать поступившее сообщение.

*Для этого используется понятие **тезаурус** пользователя.*

Тезаурус - это совокупность сведений, которыми располагает пользователь или система.

В зависимости от соотношений между **смысловым содержанием информации S** и **тезаурусом пользователя S_p** изменяется **количество семантической информации I_s** , воспринимаемой пользователем и включаемой им в дальнейшем в свой тезаурус.

Характер такой зависимости показан на рис.

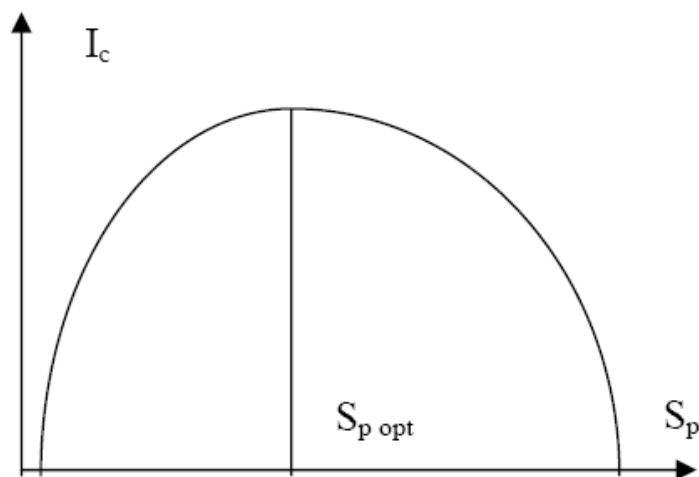


Рис 1.3. Зависимость количества семантической информации, воспринимаемой

- при $S_p=0$ пользователь не воспринимает, не понимает поступающую информацию;
- при $S_p \rightarrow \infty$ пользователь все знает, и поступающая информация ему не нужна.

Максимальное количество семантической информации I_c потребитель приобретает при согласовании ее смыслового содержания S со своим тезаурусом

S_p ($S_p = S_{p \text{ opt}}$), когда поступающая информация понятна пользователю и несет ему ранее не известные (отсутствующие в его тезаурусе) сведения.

Следовательно, количество семантической информации в сообщении, количество новых знаний, получаемых пользователем, является величиной относительной. Одно и то же сообщение может иметь смысловое содержание для компетентного пользователя и быть бессмысленным (семантический шум) для пользователя некомпетентного.

При оценке семантического (содержательного) аспекта информации необходимо стремиться к согласованию величин S и S_p .

Относительной мерой количества семантической информации может служить **коэффициент содержательности C** , который определяется как отношение количества семантической информации к ее объему:

$$C = I_c / V_d$$

4. Прагматическая мера информации

Эта мера определяет полезность информации (ценность) для достижения пользователем поставленной цели.

Эта мера также величина относительная, обусловленная особенностями использования этой информации в той или иной системе.

Ценность информации целесообразно измерять в тех же самых единицах (или близких к ним), в которых измеряется целевая функция.

Пример.

В экономической системе прагматические свойства (ценность) информации можно определить приростом экономического эффекта функционирования, достигнутым благодаря использованию этой информации для управления системой:

$$In\beta(\gamma) = \Pi(\gamma/\beta) - \Pi(\gamma),$$

где $In\beta(\gamma)$ - ценность информационного сообщения β для **системы управления γ** ;

$\Pi(\gamma)$ - априорный ожидаемый экономический эффект функционирования системы управления γ ;

$P(\gamma/\beta)$ - ожидаемый эффект функционирования системы γ при условии, что для управления будет использована информация, содержащаяся в сообщении β .

Для сопоставления введенные меры информации представим в таблице «Единицы измерения информации и примеры»

Мера информации	Единицы измерения	Примеры (для компьютерной области)
Синтаксическая : Шенноновский Подход	Степень уменьшения неопределенности	Вероятность события
Компьютерный Подход	Единицы представления информации	Бит, байт, Кбайт и т.д.
Семантическая	Тезаурус Экономические показатели	Пакет прикладных программ, персональный компьютер, компьютерные сети и т.д. Рентабельность, производительность, коэффициент амортизации и т.д.
Прагматическая	Ценность использования	Емкость памяти, произво- дительность компьютера, ско-рость передачи данных и т.д. Денежное выражение Время обработки информации и принятия решений

5. КАЧЕСТВО ИНФОРМАЦИИ

Возможность и эффективность использования информации обуславливается такими основными ее потребительскими **показателями качества**, как репрезентативность, содержательность, достаточность, актуальность, своевременность, точность, достоверность, устойчивость.

Репрезентативность информации связана с правильностью ее отбора и формирования в целях адекватного отражения свойств объекта.

Важнейшее значение здесь имеют:

- правильность концепции, на базе которой сформулировано исходное понятие;
- обоснованность отбора существенных признаков и связей отображаемого явления.

Содержательность информации отражает семантическую емкость, равную отношению количества семантической информации в сообщении к объему обрабатываемых данных, т.е. $C=I_c/V_d$.

С увеличением содержательности информации растет семантическая пропускная способность информационной системы, так как для получения одних и тех же сведений требуется преобразовать меньший объем данных.

Наряду с коэффициентом содержательности C , отражающим семантический аспект, можно использовать и **коэффициент информативности**, характеризующийся отношением количества синтаксической информации (по Шеннону) к объему данных $Y=I/V_d$.

Достаточность (полнота) информации означает, что она содержит минимальный, но достаточный для принятия правильного решения состав (набор показателей). Понятие полноты информации связано с ее смысловым содержанием (семантикой) и прагматикой.

Как неполная, т.е. недостаточная для принятия правильного решения, так и избыточная информация снижает эффективность принимаемых пользователем решений.

Доступность информации восприятию пользователя обеспечивается выполнением соответствующих процедур ее получения и преобразования. Например, в информационной системе информация преобразовывается к доступной и удобной для восприятия пользователя форме. Это достигается, в частности, и путем согласования ее семантической формы с тезаурусом пользователя.

Актуальность информации определяется степенью сохранения ценности информации для управления в момент ее использования и зависит от динамики изменения ее характеристик и от интервала времени, прошедшего с момента возникновения данной информации.

Своевременность информации означает ее поступление не позже заранее назначенного момента времени, согласованного со временем решения поставленной задачи.

Точность информации определяется степенью близости получаемой информации к реальному состоянию объекта, процесса, явления и т.п. Для информации, отображаемой цифровым кодом, известны **4 классификационных понятия точности**:

- формальная точность, измеряемая значением единицы младшего разряда;
- реальная точность, определяемая значением единицы последнего разряда числа, верность которого гарантируется;
- максимальная точность, которую можно получить в конкретных условиях функционирования системы;
- необходимая точность, определяемая функциональным назначением показателя.

Достоверность информации определяется ее свойством отражать реально существующие объекты с необходимой точностью. Измеряется достоверность информации доверительной вероятностью необходимой точности, т.е. вероятностью того, что отображаемое информацией значение параметра отличается от истинного значения этого параметра в пределах необходимой точности.

Устойчивость информации отражает ее способность реагировать на изменения исходных данных без нарушения необходимой точности. Устойчивость информации, как и репрезентативность, обусловлена выбранной методикой ее отбора и формирования.