1. Подходы к измерению количества информации. Мера Хартли.

Современная наука о свойствах информации и закономерностях информационных процессов называется теорией информации. Содержание понятия "информация" можно раскрыть на примере двух исторически первых подходов к измерению количества информации: подходов Хартли и Шеннона: первый из них основан на теории множеств и комбинаторике, а второй - на теории вероятностей.

Информация может пониматься и интерпретироваться в различных проблемах, предметных областях по-разному. Вследствие этого, имеются различные подходы к определению измерения информации и различные способы введения меры количества информации.

Количество информации - числовая величина, адекватно характеризующая актуализируемую информацию по разнообразию, сложности, структурированности (упорядоченности), определенности, выбору состояний отображаемой системы.

Если рассматривается некоторая система, которая может принимать одно из n возможных состояний, то актуальной задачей является задача оценки этого выбора, исхода. Такой оценкой может стать мера информации (события).

Мера - непрерывная действительная неотрицательная функция, определенная на множестве событий и являющаяся аддитивной (мера суммы равна сумме мер).

Меры могут быть статические и динамические, в зависимости от того, какую информацию они позволяют оценивать: статическую (не актуализированную; на самом деле оцениваются сообщения без учета ресурсов и формы актуализации) или динамическую (актуализированную т.е. оцениваются также и затраты ресурсов для актуализации информации).

Существуют различные подходы к определению количества информации. Наиболее часто используются следующие два способа измерения информации: объёмный и вероятностный.

Объёмный подход

Используется двоичная система счисления, потому что в техническом устройстве наиболее просто реализовать два противоположных физических состояния: намагничено / не намагничено, вкл./выкл., заряжено / не заряжено и др.

Объём информации, записанной двоичными знаками в памяти компьютера или на внешнем носителе информации, подсчитывается просто по количеству требуемых для такой записи двоичных символов. При этом невозмож¬но нецелое число битов.

Для удобства использования введены и более крупные, чем бит, единицы коли¬чества информации. Так, двоичное слово из восьми знаков содержит один байт информации, 1024 байта образуют килобайт (кбайт), 1024 килобайта – мегабайт (Мбайт), а 1024 мегабайта - гигабайт (Гбайт).

Энтропийный (вероятностный) подход

Этот подход принят в теории информации и кодирования. Данный способ измерения исходит из следующей модели: получатель сообщения имеет определённое представление о возможных наступлениях некоторых событий. Эти представления в общем случае недостоверны и выражаются вероятностями, с которыми он ожидает то или иное событие. Общая мера неопределённостей называется энтропией. Энтропия характеризуется некоторой математической зависимостью от совокупности вероятности наступления этих событий.

Количество информации в сообщении определяется тем, насколько уменьшилась эта мера после получения сообщения: чем больше энтропия системы, тем больше степень её неопределённости. Поступающее сообщение полностью или частично снимает эту неопределённость, следовательно, количество информации можно измерять тем, насколько понизилась энтропия системы после получения сообщения. За меру количества информации принимается та же энтропия, но с обратным знаком.

Другими, менее известными способами измерения информации являются:

Алгоритмический подход. Так как имеется много различных вычислительных машин и языков программирования, т.е. разных способов задания алгоритма, то для определённости задаётся некоторая конкретная машина, например машина Тьюринга. Тогда в качестве количественной характеристики сообщения можно взять минимальное число внутренних состояний машины, требующихся для воспроизведения данного сообщения.

Семантический подход. Для измерения смыслового содержания информации, т.е. её количества на семантическом уровне, наибольшее признание получила тезаурусная мера (тезаурус - совокупность сведений, которыми располагает пользо¬ватель или система), которая связывает семанти¬ческие свойства информации со способностью пользователя принимать поступившее сообщение.

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

В основе всей теории информации лежит открытие, сделанное Р. Хартли в 1928 году, и состоящее в том, что информация допускает количественную оценку.

Подход Р. Хартли основан на фундаментальных теоретико–множественных, по существу комбинаторных основаниях, а также нескольких интуитивно ясных и вполне очевидных предположениях.

Если существует множество элементов и осуществляется выбор одного из них, то этим самым сообщается или генерируется определенное количество информации. Эта информация состоит в том, что если до выбора не было известно, какой элемент будет выбран, то после выбора это становится известным. Необходимо найти вид функции, связывающей количество информации, получаемой при выборе некоторого элемента из множества, с количеством элементов в этом множестве, т.е. с его мощностью.

Если множество элементов, из которых осуществляется выбор, состоит из одного–единственного элемента, то ясно, что его выбор предопределен, т.е. никакой неопределенности выбора нет - нулевое количество информации.

Если множество состоит из двух элементов, то неопределенность выбора минимальна. В этом случае минимально и количество информации.

Чем больше элементов в множестве, тем больше неопределенность выбора, тем больше информации.

Количество этих чисел (элементов) в множестве равно:

N = 2i

Из этих очевидных соображений следует первое требование: информация есть монотонная функция от мощности исходного множества.

Выбор одного числа дает нам следующее количество информации:

i = Log2(N)

Таким образом, количество информации, содержащейся в двоичном числе, равно количеству двоичных разрядов в этом числе.

Это выражение и представляет собой формулу Хартли для количества информации.

При увеличении длины числа в два раза количество информации в нем также должно возрасти в два раза, несмотря на то, что количество чисел в множестве возрастает при этом по показательному закону (в квадрате, если числа двоичные), т.е. если

N2=(N1)2,

то

I2 = 2 \* I1,

F(N1\*N1)= F(N1) + F(N1).

Это невозможно, если количество информации выражается линейной функцией от количества элементов в множестве. Но известна функция, обладающая именно таким свойством: это Log:

Log2(N2) = Log2(N1)2= 2 \* Log2(N1).

Это второе требование называется требованием аддитивности.

Таким образом, логарифмическая мера информации, предложенная Хартли, одновременно удовлетворяет условиям монотонности и аддитивности. Сам Хартли пришел к своей мере на основе эвристических соображений, подобных только что изложенным, но в настоящее время строго доказано, что логарифмическая мера для количества информации однозначно следует из этих двух постулированных им условий.

Пример. Имеются 192 монеты. Известно, что одна из них - фальшивая, например, более легкая по весу. Определим, сколько взвешиваний нужно произвести, чтобы выявить ее. Если положить на весы равное количество монет, то получим 3 независимые возможности: а) левая чашка ниже; б) правая чашка ниже; в) чашки уравновешены. Таким образом, каждое взвешивание дает количество информации I=log23, следовательно, для определения фальшивой монеты нужно сделать не менее k взвешиваний, где наименьшее k удовлетворяет условию log23k log2192. Отсюда, k 5 или, k=4 (или k=5 - если считать за одно взвешивание и последнее, очевидное для определения монеты). Итак, необходимо сделать не менее 5 взвешиваний (достаточно 5).

2. Социальные аспекты информатики.

Термин “социальные аспекты” применительно к большей части наук, тем более фундаментальных, звучит странно. Вряд ли фраза “Социальные аспекты математики” имеет смысл. Однако, информатика – не только наука.

И впрямь, мало какие факторы так влияют на социальную сферу обществ (разумеется, находящихся в состоянии относительно спокойного развития, без войн и катаклизмов) как информатизация. Информатизация общества – процесс проникновения информационных технологий во все сферы жизни и деятельности общества. Многие социологи и политологи полагают, что мир стоит на пороге информационного общества. В. А. Извозчиков предлагает следующее определение: “Будем понимать под термином “информационное” (“компьютеризированное”) общество то, во все сферы жизни и деятельности членов которого включены компьютер, телематика, другие средства информатики в качестве орудий интеллектуального труда, открывающих широкий доступ к сокровищам библиотек, позволяющих с огромной скоростью проводить вычисления и перерабатывать любую информацию, моделировать реальные и прогнозируемые события, процессы, явления, управлять производством, автоматизировать обучение и т.д.”. Под “телематикой” понимаются службы обработки информации на расстоянии (кроме традиционных телефона и телеграфа).

Последние полвека информатизация является одной из причин перетока людей из сферы прямого материального производства в, так называемую, информационную сферу. Промышленные рабочие и крестьяне, составлявшие в середине XX века более 2/3 населения, сегодня в развитых странах составляют менее 1/3. Все больше тех, кого называют “белые воротнички” – людей, не создающих материальные ценности непосредственно, а занятых обработкой информации (в самом широком смысле): это и учителя, и банковские служащие, и программисты, и многие другие категории работников. Появились и новые пограничные специальности. Можно ли назвать рабочим программиста, разрабатывающего программы для станков с числовым программным управлением? – По ряду параметров можно, однако его труд не физический, а интеллектуальный.

Информатизация сильнейшим образом влияет на структуру экономики ведущих в экономическом отношении стран. В числе их лидирующих отраслей промышленности традиционные добывающие и обрабатывающие отрасли оттеснены максимально наукоемкими производствами электроники, средств связи и вычислительной техники (так называемой, сферой высоких технологий). В этих странах постоянно растут капиталовложения в научные исследования, включая фундаментальные науки. Темпы развития сферы высоких технологий и уровень прибылей в ней превышают в 5-10 раз темпы развития традиционных отраслей производства. Такая политика имеет и социальные последствия – увеличение потребности в высокообразованных специалистах и связанный с этим прогресс системы высшего образования. Информатизация меняет и облик традиционных отраслей промышленности и сельского хозяйства. Промышленные роботы, управляемые ЭВМ, станки с ЧПУ стали обычным оборудованием. Новейшие технологии в сельскохозяйственном производстве не только увеличивают производительность труда, но и облегчают его, вовлекают более образованных людей.

Казалось бы, компьютеризация и информационные технологии несут в мир одну лишь благодать, но социальная сфера столь сложна, что последствия любого, даже гораздо менее глобального процесса, редко бывают однозначными. Рассмотрим, например, такие социальные последствия информатизации как рост производительности труда, интенсификацию труда, изменение условий труда. Все это, с одной стороны, улучшает условия жизни многих людей, повышает степень материального и интеллектуального комфорта, стимулирует рост числа высокообразованных людей, а с другой – является источником повышенной социальной напряженности. Например, появление на производстве промышленных роботов ведет к полному изменению технологии, которая перестает быть ориентированной на человека. Тем самым меняется номенклатура профессий. Значительная часть людей вынуждена менять либо специальность, либо место работы – рост миграции населения характерен для большинства развитых стран. Государство и частные фирмы поддерживают систему повышения квалификации и переподготовки, но не все люди справляются с сопутствующим стрессом. Прогрессом информатики порожден и другой достаточно опасный для демократического общества процесс – все большее количество данных о каждом гражданине сосредоточивается в разных (государственных и негосударственных) банках данных. Это и данные о профессиональной карьере (базы данных отделов кадров), здоровье (базы данных учреждений здравоохранения), имущественных возможностях (базы данных страховых компаний), перемещении по миру и т.д. (не говоря уже о тех, которые копят специальные службы). В каждом конкретном случае создание банка может быть оправдано, но в результате возникает система невиданной раньше ни в одном тоталитарном обществе прозрачности личности, чреватой возможным вмешательством государства или злоумышленников в частную жизнь. Одним словом, жизнь в “информационном обществе” легче, по-видимому, не становится, а вот то, что она значительно меняется – несомненно.