

Задача об измерении информации

Т.Ю. Селихова

Автор: Селихова Т.Ю., учитель информатики и ИКТ средней школы № 4 с. Монастырище Черниговского района Приморского края.

Предмет: Информатика и ИКТ.

Класс: 10.

Тема: Измерение информации.

Профиль: Общеобразовательный.

Уровень: Продвинутый.

Текст задачи. Проблема определения количества информации решается учёными до сих пор. Известно несколько подходов к измерению количества информации: структурный, статистический и семантический.

Почему нельзя обойтись одним способом измерения информации и какой из них наиболее правильный?

а) Выделите ключевые слова для информационного поиска.

б) Найдите и соберите необходимую информацию.

в) Обсудите и проанализируйте собранную информацию.

г) Сделайте выводы.

д) Сравните ваши выводы с культурным образцом.

Возможные информационные источники

Web-сайты:

http://www.cko-bait.ru/index.php?option=com_content&view=article&id=216%3A2009-05-07-06-44-03&catid=79%3A2009-05-07-06-30-43&Itemid=92&lang=

<http://www.metodolog.ru/00262/00262.html>

<http://letopisi.ru/index.php>

Культурный образец

Бекман И.Н. Компьютерные науки. Курс лекций. Лекция 3. Информация. <http://profbeckman.narod.ru/Komp.files/Lec3.pdf>

Информационный объём сообщения (информационная ёмкость сообщения) — количество информации в сообщении, измеренное в битах, байтах или производных единицах (Кбайтах, Мбайтах и т.д.).

Количество информации — мера уменьшения неопределённости.

Количество технической информации — числовая характеристика сигнала, которая не зависит от его формы и содержания и характеризует неопределённость, которая исчезает после получения сообщения в виде данного сигнала. Оно зависит от вероятности получения сообщения о том или ином событии. Для абсолютно достоверного события (событие обязательно произойдёт, поэ-

тому его вероятность равна 1) количество информации в сообщении о нём равно 0. Чем невероятнее событие, тем большее количество информации несёт сообщение о нём. Лишь при равновероятных ответах ответ «да» или «нет» несёт один бит информации.

...В компьютерной технике измерению обычно подвергается информация, представленная дискретным сигналом. При этом различают следующие подходы.

Структурный (алфавитный, объёмный). Измеряет количество информации простым подсчётом информационных элементов, составляющих сообщение. Применяется для оценки возможностей запоминающих устройств, объёмов передаваемых сообщений, инструментов кодирования без учёта статистических характеристик их эксплуатации. Алфавитный подход к измерению информации не связывает количество информации с содержанием сообщения. Это — объективный подход к измерению информации. Количество информации зависит от объёма текста и мощности алфавита. Ограничений на максимальную мощность алфавита нет, но есть достаточный алфавит мощностью 256 символов. Этот алфавит используется для представления текстов в компьютере. Поскольку $256=2^8$, то один символ несёт в тексте 8 бит информации.

Статистический (вероятностный). Учитывает вероятность появления сообщений: более информативным считается то сообщение, которое менее вероятно, т.е. менее всего ожидалось. Применяется при оценке значимости получаемой информации. Все события происходят с раз-

личной вероятностью, но зависимость между вероятностью событий и количеством информации, полученной при совершении того или иного события можно выразить формулой Шеннона.

...*Семантический* (содержательный). Учитывает целесообразность и полезность информации. Применяется при оценке эффективности получаемой информации и её соответствия реальности. Сообщение — информативный поток, который в процессе передачи информации поступает к приёмнику. Сообщение несёт информацию для человека, если содержащиеся в нём сведения являются для него новыми и понятными. Информация — знания человека — сообщение должно быть информативно. Если сообщение не информативно, то количество информации с точки зрения человека = 0. (Пример: вузовский учебник по высшей математике содержит знания, но они не доступны первокласснику).

...В рамках алфавитного (структурного) подхода выделяют три меры информации.

Геометрическая. Определяет максимально возможное количество информации в заданных объёмах.

Мера может быть использована для определения информационной ёмкости памяти компьютера.

...Очевидно, геометрическая мера не учитывает, какими символами заполнено сообщение. Так, одинаковыми по количеству информации, измеренной геометрической мерой, являются, например, сообщения «компьютер» и «программа»; а также 346 и 10В.

Комбинаторная. Оценивает возможность представления инфор-

РЕСУРСЫ

мации при помощи различных комбинаций информационных элементов в заданном объёме.

...Комбинаторная мера может использоваться для оценки информационных возможностей некоторой системы кодирования.

...Использует типы комбинаций элементов и соответствующие математические соотношения, которые приводятся в одном из разделов дискретной математики — комбинаторике.

...Комбинаторная мера может использоваться для оценки информационных возможностей некоторого автомата, который способен генерировать дискретные сигналы (сообщения) в соответствии с определённым правилом комбинаторики.

Комбинаторная мера используется для определения возможностей кодирующих систем.

...*Аддитивная (мера Хартли).* Хартли рассмотрел кодировку сообщения с помощью некоторого набора знаков (если для данного набора установлен порядок следования знаков, то он называется алфавитом). Самой сложной частью работы оказалось определение количества информации, содержащейся в каждом отдельном символе.

...Алфавит — вся совокупность символов, используемых в некотором языке для представления информации.

...С целью анализа некоторого текста введём обозначения: N — мощность алфавита, использованного для написания текста, т.е. число символов в алфавите (размер алфавита); n — число символов в сообщении (длина текста, например, одного слова); I — количество информации в

сообщении; $i=l/n$ — информационный вес символа (количество информации в одном символе).

Полная информация, содержащаяся в сообщении, определяется как количество сведений (при их длине n) пропорциональное числу смысловых символов N формулой Хартли: $I = n \log_a N$

Согласно этому соотношению, количество информации в передаваемом сообщении пропорционально его длительности (числу символов). Выбор основания логарифма a влияет только на размерность, т.е. на единицу измерения количества информации. Наиболее удобным оказалось основание логарифма $a = 2$.

...При алфавитном подходе к измерению информации количество информации зависит не от содержания, а от размера текста и мощности алфавита.

Применение алфавитного подхода удобно при использовании технических средств работы с информацией. Алфавитный подход является объективным способом измерения информации в отличие от субъективного содержательного подхода. Удобнее всего измерять информацию, когда размер алфавита N равен целой степени двойки.

...Другой подход к измерению информации — статистический (вероятностный) — рассматривает информацию как снятую неопределённость. К. Шеннон предложил связать количество информации, которое несёт в себе некоторое сообщение, с вероятностью получения этого сообщения.

...Шеннон измерял количество информации как меру достоверности передаваемого сигнала в битах.

Информация уничтожает неопределённость. Степень неопределённости принято характеризовать с помощью понятия «вероятность».

Вероятность — величина, которая может принимать значения в диапазоне от 0 до 1. Она может рассматриваться как мера возможности наступления какого-либо события, которое может иметь место в одних случаях и не иметь места в других.

...Для определения информации в одном символе алфавита можно также использовать вероятностные методы, поскольку появление конкретного знака в конкретном месте текста есть явление случайное.

Формула Шеннона:

$$I = \sum_{j=1}^n p_j \log_2 \frac{1}{p_j}$$

где I — символ количества информации, воплощённой в некоторой системе, j — индекс состояний системы, n — число состояний, p_j — вероятность состояния j .

Формула Шеннона предназначена для измерения количества информации в системах, которым присуще конечное количество дискретных состояний, различающихся по распространённости внутри соответствующих систем.

...Доказано, что эта формула выражает единственно возможную меру количества информации в системах указанного в ней типа (с точностью до постоянного множителя, который служит для выбора единицы информации). Величина I в формуле Шеннона представляет собой математическое ожидание информации, воп-

лощённой в некоторой системе, имеющей различные состояния j .

...Сравним вероятностный и алфавитный подходы. Первый подход позволяет вычислить предельное (минимально возможное) теоретическое значение количества информации, которое несёт сообщение о данном исходе события. Второй — каково количество информации на практике с учётом конкретной выбранной кодировки.

Очевидно, что первая величина есть однозначная характеристика рассматриваемого события, тогда как вторая зависит ещё и от способа кодирования: в «идеальном» случае обе величины совпадают, однако на практике используемый метод кодирования может иметь ту или иную степень избыточности.

С рассмотренной точки зрения вероятностный подход имеет преимущество. Но, с другой стороны, алфавитный способ заметно проще и с некоторых позиций (например, для подсчёта требуемого количества памяти) полезнее.

Методический комментарий

Задача относится к профильному обучению. Теоретический материал, необходимый для решения задачи, обобщает и углубляет знания учащихся о способах измерения информации. В культурном образце не звучит однозначный ответ на поставленный вопрос. Вывод и обоснование предлагается сделать самостоятельно.

Ключевыми являются фразы: «измерение информации», «структурный подход», «статистический подход», «семантический подход».