

## ИНФОРМАЦИОННЫЙ АНАЛИЗ ИГРЫ «БЫКИ И КОРОВЫ»

От редакции: статья посвящена обсуждению задачи «Быки и коровы» конкурса КИО-2007.

### КАК ВЫЧИСЛИТЬ КОЛИЧЕСТВО ИНФОРМАЦИИ

Каждый из нас постоянно слышит и употребляет в повседневной речи высказывания такого рода:

– «за последние годы было создано столько информации, сколько за всю предыдущую историю человечества»,

– «в этой книге много информации на тему...»,

– «на этом диске записан большой объем аудио и видео информации».

Что же при этом мы понимаем под количеством передаваемой информации?

Ведь на диск мы можем записать бессмысленную какофонию звуков, а книга может быть на неизвестном нам языке. Вряд ли при этом мы получим какую-либо информацию. В замечательной статье С.С. Лаврова «Так что же такое информатика?» (КИО № 3–4, 1999) писалось о том, что, к сожалению, в обиход вместо термина «данные» вошел термин «информация», и, если с передачей или вычислением объема данных все понятно, то передача информации во многом определяется тем, кто ее получает. И действительно, объем данных в книге на китайском языке мы можем считать как число иероглифов или как число двоичных разрядов в бинарной кодировке текста книги. В то же время, объем информации, полученный от ее чтения человеком, не знающим китайского языка, будет близок к нулю (ка-

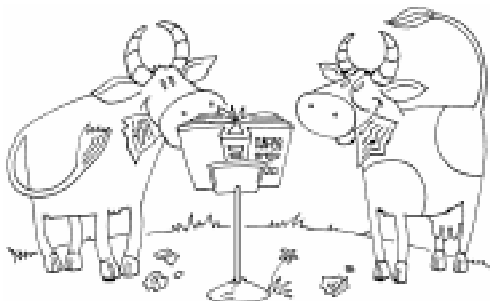
кую-то информацию можно получить, просто анализируя текст книги без понимания его смысла, как пытаются анализировать сигналы из космоса).

Итак, получение информации человеком связано с теми изменениями, которые происходят в мозге при получении тех или иных данных. Для тех, кто уже знает смысл передаваемого сообщения либо не способен осознать этот смысл, количество полученной информации будет равным нулю. Для тех же, кто осмыслит получаемые данные, они превратятся в полученную информацию.

Познакомимся с простой моделью измерения информации, рассматривая ее как меру уменьшения неопределенности.

Наиболее понятно рассмотреть эту модель на примере игры «Угадай-ка!». В этой игре один из участников задумывает некоторое число от 1 до 1024, а другой должен угадать его за наименьшее число вопросов, сформулированных так, чтобы на них можно было отвечать «да» или «нет». Например, на вопрос «есть ли в этом числе ноль» можно получить только ответы «да» или «нет». При ответе «да» количество возможных чисел уменьшится на 825 (до 199), а при ответе «нет» только на 199 (до 825). Ясно, что в первом случае получено больше информации, чем во втором.

Чтобы измерить количество информации, полученной в каждом случае, введем понятие бита – количества информации, необходимо-



*...передача информации  
во многом определяется тем,  
кто её получает.*

го для уменьшения числа возможных случаев (неопределенности) вдвое. Таким образом, вопрос «число больше 512?» при любом ответе уменьшает число подозрительных чисел вдвое и значит, задавая этот вопрос, мы получаем 1 бит информации при любом ответе. Действуя по принципу половинного деления, мы сможем угадать число за десять вопросов. Таким образом, можно считать, что по отношению к этой модели исходная задача имеет информационную ёмкость в 10 бит.

Что будет, если, например, один игрок загадал число 123, а другой задал вопрос «это число 123?». В этом случае число угадано сразу, а, значит, за один ход получено 10 бит информации. При этом количество подозрительных слов уменьшилось в 1024 раза. Поскольку  $1024 = 2^{10}$ , этот ответ можно записать как  $\log_2 1024 = 10$ .

Итак, мы можем определить количество полученной информации как двоичный логарифм от числа раз, в которое уменьшилась неопределенность после получения конкретного ответа на заданный вопрос.

Возвращаясь к вопросу первого примера, теперь можно оценить количество полученной информации при разных ответах:

– при ответе «да» число «подозрительных» чисел уменьшается от 1024 до 199, и мы получаем количество информации  $\log_2 1024/199 = 2,36...$  бит,

– в случае отрицательного ответа количество информации меньше, а именно,  $\log_2 1024/825 = 0,31...$  бит. Таким образом, количество информации в первом случае примерно на 2 бита больше, чем во втором.

**Теорема.** *В описанной выше задаче при любых вопросах и ответах на них («да» или «нет») количество полученной информации в результате будет одинаково.*

**Доказательство**

Пусть исходно нам надо было угадать одно из  $N$  чисел. Допустим, в процессе угадывания количество возможных слов уменьшалось (точнее не увеличивалось) в такой последовательности:

$$n_1 = N, n_2, n_3, \dots, n_k = 1.$$

Тогда общее количество информации, полученное в процессе угадывания, будет:

$$\begin{aligned} & \log_2 (n_1/n_2) + \log_2 (n_2/n_3) + \dots + \log_2 (n_{k-1}/n_k) \\ & = \log_2 [(n_1/n_2)(n_2/n_3)\dots(n_{k-1}/n_k)] = \\ & = \log_2 (n_1/n_k) = \log_2 N \end{aligned}$$

В нашей задаче это количество, как уже говорилось выше, равно  $\log_2 1024 = 10$ .

Заметим, что бывают задачи, в которых за один вопрос можно получить информации больше, чем 1 бит. Например, в известной задаче о взвешиваниях имеется 27 монет, одна из которых фальшивая и легче остальных. Нужно, используя рычажные весы (на которых можно только сравнить два груза и узнать, который весит больше), найти фальшивую монету за наименьшее число взвешиваний. Если разделить монеты на три равные части (по 9 монет в каждой) и сравнить на весах две из них, то мы сразу уменьшим число подозреваемых монет втрое (если весы показывают равновесие, то фальшивая монета в оставшейся кучке). Таким образом, за три взвешивания мы найдем фальшивую монету. За одно взвешивание (при любом раскладе) мы будем получать  $\log_2 3 = 1,58...$  бит информации. Значит информационный объем нашей задачи равен  $\log_2 27 = 4,75...$  бит.

*Замечание.* В этой задаче удобнее считать в «тритах», определяя их как информацию, необходимую для уменьшения неопределенности в три раза. Тогда информационная ёмкость задачи будет равна 3 тритам. Один трит равен  $\log_2 3$  бита.

**ИНФОРМАЦИОННЫЙ АНАЛИЗ  
ИГРЫ «БЫКИ И КОРОВЫ»**

В этом разделе будет рассмотрен один из сюжетов конкурса КИО-2007, построенный на известной словесной игре «Быки и коровы». Игра «Быки и коровы» состоит в следующем.

Один из игроков задумывает слово (в нашей задаче это будет трехбуквенное слово из предложенного набора, в котором 170 слов). Второй игрок пытается отгадать задуманное слово. Для этого он называет любые слова из того же числа букв (в нашей



Рис. 1

задаче слова из базового набора). Первый игрок в ответ на каждое слово сообщает, какое количество букв в названном и задуманном слове совпадает и находится на тех же местах (число «быков») и какое количество букв совпадает, но находится на разных местах (число «коров»).

Например, если 1-ый игрок задумал слово «рот», а второй сказал «ток», то ответ будет «1 бык и 1 корова».

Учитывая сказанное выше, мы можем подсчитать информационный объем задачи:  $\log_2 170 = 7,4...$  Представляет интерес, выяснить какую наибольшую информацию можно получить при задании одного вопроса в этой игре, какова средняя информативность правильно задаваемых вопросов (то есть ведущих к наименьшему числу ходов), можно ли подобрать такую стратегию из оптимального числа ходов, при которой на каждом ходе добывается примерно одинаковая информация.

В предложенной участникам среде для поддержки экспериментов роль игрока, «задумывающего» слово, играет компьютер, но играет компьютер «нечестно»: на каждом ходе он «перезадумывает» слово, следя за тем, чтобы это изменение сохраняло пра-

вильными все предыдущие ответы. Это сделано для того, чтобы анализ игры не зависел от случая, и стратегия, которую в результате находит участник-экспериментатор, годилась для любого задуманного слова.

Целью эксперимента было найти такую стратегию, которая за минимальное число шагов сузила бы число слов, из которых компьютер «перезадумывает» слово, до единственно возможного.

Найденная стратегия оценивается также по дополнительным признакам. Они связаны с количеством извлекаемой на каждом ходе информации.

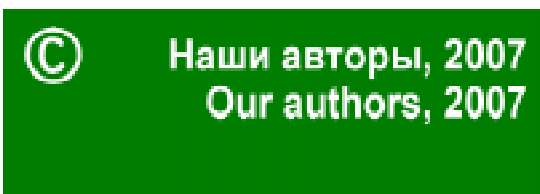
Отмечаются стратегии:

- 1) с лучшим ходом (ходом, с помощью которого извлекается максимальное количество информации);
- 2) самые стабильные стратегии, то есть стратегии, при которых на каждом ходе добывается примерно равное количество информации (среднеквадратичное отклонение от среднего минимально).

На рис. 1 приведены два решения с минимальным числом ходов. При правильной игре за один ход можно получить в среднем 1,235 бит информации. Самый результативный ход приносит 2,585 бит, однако сыграть так, чтобы за каждый ход получать одинаковое число информации не получается: среднеквадратичное отклонение от среднего составляет не менее 1,497 бита.

Эти результаты были получены экспериментально в ходе конкурса КИО-2007, однако дать им теоретическое обоснование трудно вследствие разнообразия используемых слов-тестеров (это исходный список из 170 слов).

Попробуйте применить предложенный анализ к другим играм-угадайкам.



Поздняков Сергей Николаевич,  
профессор кафедры ВМ-2  
СПбГЭТУ (ЛЭТИ).