**Введение в теорию информации**

Когда я учился в институте, очень популярной была игра «[Быки и коровы](http://ru.wikipedia.org/wiki/%D0%91%D1%8B%D0%BA%D0%B8_%D0%B8_%D0%BA%D0%BE%D1%80%D0%BE%D0%B2%D1%8B)». Я был практически непобедимым игроком, а всё благодаря относительно несложному алгоритму, разработанному мною на основе теории информации. И вот недавно я обнаружил, что в любимую игру моей юности можно сразиться [онлайн](http://www.oru.ru/games/gamesonline/_games/igra/495.html). И что вы думаете? Я сыграл трижды против самого сильного «соперника» (из тех, что предложил компьютер), и все три раза выиграл! ☺

Я решил рассказать о разработанном мною алгоритме, но для начала – о книге, которая познакомила меня с основами теории информации. В печатном виде найти ее непросто (всё же год издания 1980-й), а вот в электронном – пожалуйста.

Альфред Реньи. Трилогия о математике. – М.: Мир, 1980. – 376 с. В сборник включены научно-популярные произведения известного венгерского математика и популяризатора науки: «Диалоги о математике», «Письма о вероятности», «Дневник. — Записки студента по теории информации», а также четыре статьи: о теории вероятностей, о ее преподавании, о числах Фибоначчи и о математической теории «деревьев».



Книга интересна вся, но к разработке алгоритма игры «Быки и коровы» «причастна» лишь новелла «Дневник. — Записки студента по теории информации». Краткое её изложение я вам и представляю.

Возникновение математической теории информации стало возможным после того, как было осознано, что количество информации можно задать числом. Чтобы измерить количество любой информации, необходимо закодировать ее в виде последовательности нулей и единиц наиболее рациональным способом (то есть способом, позволяющим получить самую короткую последовательность). Длина (количество знаков) полученной последовательности нулей и единиц (кодового слова) может служить мерой количества информации. Ответ на вопрос, допускающий лишь два ответа—«да» и нет», а в остальном произвольный, служит единицей для измерения количества информации.

Каждая цифра числа, записанного в двоичной системе счисления, содержит одну единицу информации. Отсюда и название единицы информации — бит (от английского *bi*nary digi*t* —двоичная цифра).

Нa лекции присутствовало 32 студента. Лектор спросил, сколько вопросов необходимо задать, чтобы установить, кого из нас он задумал. Я тотчас ответил, что для этого понадобится 5 вопросов. Тогда лектор попросил меня пояснить, какие вопросы я имею в виду. Я сказал, что, прежде всего, составил бы список присутствующих, расположив их фамилии в алфавитном порядке, и спросил бы (в этом и состоял бы мой первый вопрос), не внесен ли задуманный лектором студент в первую половину списка. Полученный ответ, каким бы он ни был — утвердительным или отрицательным, позволил бы сузить круг поисков, вдвое уменьшив число «подозреваемых» с 32 до 16. Действуя аналогично, я вторым вопросом сократил бы число студентов, среди которых может оказаться задуманный лектором слушатель, до 8, третьим вопросом — до 4, четвертым — до 2 и, задав пятый вопрос, установил бы, кто был задуман. Затем лектор поинтересовался, хватило бы мне пяти вопросов, если бы их надо было задавать все сразу, то есть если бы, ставя очередной вопрос, я не знал ответов на предыдущие вопросы. Мне казалось, что если все вопросы задавать сразу, то их понадобится больше пяти. Но я заблуждался. Лектор показал, как следует задать сразу пять вопросов, чтобы по ответам на них можно было установить, кто из студентов был задуман. Перенумеруем сидящих в аудитории последовательными целыми числами от 0 до 31 и запишем порядковый номер каждого студента в двоичной системе. Каждой из 32 фамилий мы тем самым поставим в соответствие пятиразрядное двоичное число (если двоичная запись порядкового номера окажется короче пяти знаков, припишем недостающие знаки (нули) слева; например, число 1 запишем в виде 00001. Пять вопросов, которые можно задать сразу, заключаются в следующем: верно ли, что в двоичной записи порядкового номера фамилии задуманного студента первая, вторая, третья, четвертая и пятая цифры равны единице?

*Если в заданном множестве* ***Н****, содержащем* ***N*** *элементов, выделен какой-то элемент* ***х****, о котором заранее известно лишь, что он принадлежит множеству* ***Н****, то, чтобы найти* ***х****, необходимо получить количество информации, равное* ***log2N*** *битам. Эту формулу обычно называют формулой Хартли.*

Закон аддитивности информации:

(1) log2N1N2 = log2N1 + log2N2

При **неудачном выборе вопросов** новой является лишь какая-то доля информации, а другая ее часть содержится в информации, полученной при ответе на предыдущие вопросы.

Формула Хартли исходит из молчаливого предположения о том, что все ***N*** элементов равновероятны. Однако в действительности такое предположение выполняется редко.

В общем же случае количество информации Н(ξ) можно найти по формуле Шеннона:

(2) Н(ξ) = p1log2(1/p1) + p2log2(1/p2) + … + pNlog2(1/pN),

где ξ – неизвестный элемент множества Н = {х1, х2, …, хN}, то есть ξ – случайная величина, принимающая значения х1, х2, …, хN, а рK – вероятность того, что ξ принимает значения хK (k = 1, 2, … N).

По возможности всегда следует стремиться к тому, чтобы утвердительные и отрицательные ответы на задаваемые нами вопросы были равновероятны.

Вероятность наступления независимых событий равна произведению их вероятностей.

Если при фиксированном значении **N** распределение вероятностей (p1, р2, ..., рN) отлично от ***равномерного*** распределения, то сумма p1log2(1/p1) + p2log2(1/p2) + … + pNlog2(1/pN) всегда меньше чем log2N. Это означает следующее: если о случайной величине ξ известно лишь, что она принимает N различных значений, то при наблюдении одного значения ξ наибольшую информацию мы получаем в том случае, если все возможные значения ξ равновероятны.

Формулу (2) Винер вывел в 1948 г. независимо от Шеннона. Однако еще в XIX веке эта формула встречалась в работах Больцмана, поэтому ее часто называют формулой Больцмана — Шеннона. Больцман пришел к ней, занимаясь совсем другой задачей – изучением энтропии. Энтропия физической системы служит мерой неупорядоченности системы. Следовательно, можно считать, что энтропия системы является мерой неопределенности состояния молекул, образующих систему.

Неопределенность есть не что иное, как недостача информации, или отрицательная информация. Ту же мысль можно выразить иначе: информация есть не что иное, как убыль неопределенности. До наблюдения случайной величины ξ мы пребываем в полной неопределенности относительно того, какое из своих значений она может принять. После того как наблюдение произведено, неопределенность устраняется. Наблюдение случайной величины ξ содержит Н(ξ) битов информации; поскольку эта информация позволила избавиться от неопределенности, существовавшей до наблюдения, естественно принять число Н(ξ) за меру этой неопределенности. Таким образом, Н(ξ) можно рассматривать относительно значения, принимаемого величиной ξ, и как меру неопределенности, существовавшей до наблюдения случайной величины ξ. Эту меру неопределенности и принято называть энтропией. Итак, формула Шеннона допускает следующую интерпретацию. Если случайная величина ξ принимает значения х1, х2, …, хN с вероятностями р1, р2, …, рN, то энтропию величины ξ (то есть меры неопределенности существовавшей до наблюдения величины ξ), которую мы обозначим Н(ξ), можно вычислить по формуле:

(3) Н(ξ) = $\sum\_{k=1}^{N}p$klog2(1/pk)

Кодовый знак, который может принимать лишь два значения, содержит один бит информации только тогда, когда оба его значения равновероятны (то есть вероятность каждого из них равна ½), а во всех остальных случаях количество информации, содержащейся в таком знаке, меньше одного бита (рис. 1).



Рис. Зависимость количества информации (энтропии) от вероятности одного из двух событий

Количество информации о случайной величине ξ, полученной при наблюдении случайной величины η:



Относительно величины I(ξ, η) можно утверждать следующее:

а) Величина I(ξ, η) всегда неотрицательна и обращается в 0 лишь в том случае, если ξ и η независимы. Если ξ и η независимы, то наблюдение случайной величины η; не дает никакого выигрыша в информации относительно случайной величины ξ.

б) Справедливо неравенство I(ξ, η) ≤ Н(ξ), причем равенство в нем достигается в том и только в том случае, если ξ = f(η), то есть если значения случайной величины η однозначно определяют значения случайной величины ξ. Действительно, в этом случае наблюдение случайной величины η позволяет точно установить значение случайной величины ξ, то есть получить полную информацию относительно ξ. Иначе говоря, наблюдение случайной величины η позволяет полностью устранить всю неопределенность Н(ξ), относительно случайной величины ξ.

в) Величина I(ξ, η) симметрична: I(ξ, η) = I(η, ξ), то есть наблюдение случайной величины η позволяет получить точно такое же количество информации относительно случайной величины ξ, какое наблюдение случайной величины ξ дает относительно случайной величины η. Именно поэтому величину I(ξ, η) принято называть взаимной информацией случайных величин ξ и η. Равенство I(ξ, η) = I(η, ξ) также означает, что если мы рассматриваем две случайные величины, в определенной мере зависящие друг от друга, то средства теории информации не позволяют определить, какая из них играет роль причины и какая — следствия. Единственное, что в наших силах, — это установить, насколько тесна связь между двумя случайными величинами – ***насколько велика корреляция!***