ИНФОРМАЦИЯ

10 класс

Учитель Малахова Е.В.

2010 год

Измерение количества информации

Вставьте пропущенные слова:

- Сообщение называется ……информативным……, если в нём содержатся новые и понятные сведения.

- 1 бит - это количество информации, ……уменьшающее……. неопределенность знаний в 2 раза.

- События, не имеющие преимущество друг перед другом, называются ……равновероятностными…….

*

Измерение количества информации

Ответьте на вопросы:

(Ответ: 3 бита)

1. Какой объем информации содержит сообщение, уменьшающее неопределенность знаний в 8 раз?
2. Чему равна информационная емкость знака генетического алфавита ?

(Ответ: 2 бита)

3. В доме 16 этажей. На каждом этаже по несколько квартир. Сообщение о том, что Саша живёт в квартире № 40, содержит 6 бит информации. Сколько квартир на каждом этаже?

64 квартиры, значит
на каждом этаже по 4 квартиры

4. Найти объем инфор
«Встреча назначена на завтра.»

28 байт или
28*8=224 бита

Измерение количества информации

Какое сообщение содержит большее количество информации?

(Ответ: 2 бита)

1. Вася получил за контрольную работу оценку 4 (по 5-бальной системе единицы не ставят).

(Ответ: 2 бита)

2. По информатике Вася имеет в журнале 3 пятерки, 3 четверки , 3 тройки и 3 двойки. Вася получил оценку «5».

3. За контрольную работу по информатике получено 10 пятерок, 13 четверок, 5 троек и 2 двойки. Вася получил тетрадь с оценкой?

?

*

ИНФОРМАЦИЯ

Вероятностный подход к определению количества информации.

Вероятностный подход

Цель урока:

- сформировать понимание вероятности;
  - научиться находить информационный объем неравновероятностных событий;
  - установить связь между алфавитным и содержательным подходом к определению количества информации.

*

Вероятностный подход

Вероятность события выражается в долях единицы и равна отношению количества повторений данного исхода события к общему числу повторений события:

$$p = K/N, \text{ где}$$

К – величина, показывающая сколько раз произошло интересующее нас событие,

N – общее число возможных исходов какого-то процесса.

p = 1 событие происходит **всегда** (нет неопределенности)

p = 0,5 событие происходит в половине случаев (есть неопределенность)

p = 0 событие **никогда** не происходит (нет неопределенности)

*

Вычисление вероятности

Задача. За контрольную работу по информатике получено 10 пятерок, 13 четверок, 5 троек и 2 двойки. Какое количество информации получил Вася при получении тетради с оценкой?

Формула: $p_i = \frac{n_i}{N}$

число «нужных» событий

общее число событий

Решение: $N = K5+K4+K3+K2=10+13+5+2=30$

вероятность «5»: p_5 =10/30=0,33

вероятность «4»: p_4 =13/30=0,43

вероятность «3»: p_3 =5/30=0,17

вероятность «2»: p_2 =2/30=0,07

Как иначе посчитать *P2*?

$$p_2 = 1 - p_5 - p_4 - p_3 = 1-0,33-0,43-0,17=0,07$$

*

Вычисление количества информации

Формула: $2^I = \frac{1}{p}$ или $I_i = -\log_2 p_i = \log_2 \frac{1}{p_i}$

Решение:

Логарифмом числа b по основанию a называют такое число x, что $a^x = b$. При этом пишут:

$$x = \log_a b$$

$$I_3 = -\log_2 0{,}17 = \log_2\left(\frac{1}{0{,}17}\right) = \ldots\ldots\ldots\ldots\ldots\ldots бит$$

$$I_2 = -\log_2 0{,}07 = \log_2\left(\frac{1}{0{,}07}\right) = \underline{3{,}836}\ldots\ldots\ldots бит$$

*

Вероятностный подход

Как посчитать количество информации для любого неравновероятностного события?

Клод Элвуд Шеннон (1916—2001) американский математик и электротехник, один из создателей математической теории информации и криптографии.

$$I = -(p_1 \cdot \log_2 p_1 + p_2 \cdot \log_2 p_2 + ... + p_n \cdot \log_2 p_n) = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i$$

где I - количество информации;

n - количество возможных событий;

p_i - вероятность i-го события.

*

Вероятностный подход

Количество информации о получении любой из 4 оценок:

$$I = -(p_5 \cdot \log_2 p_5 + p_4 \cdot \log_2 p_4 + p_3 \cdot \log_2 p_3 + p_2 \cdot \log_2 p_2) =$$

$$-(0{,}33 \cdot \log_2 0{,}33 + 0{,}43 \cdot \log_2 0{,}43 + 0{,}16 \cdot \log_2 0{,}16 + 0{,}07 \cdot \log_2 0{,}07) \approx$$

$$\boxed{\approx 1{,}74 \text{ бит}}$$

Измерение количества информации

Какое сообщение содержит большее количество информации?

(Ответ: 2 бита)

1. Вася получил за контрольную работу оценку 4 (по 5-бальной системе единицы не ставят).

(Ответ: 2 бита)

2. По информатике Вася имеет в журнале 3 пятерки, 3 четверки , 3 тройки и 3 двойки. Вася получил оценку «5».

3. За контрольную работу по информатике получено 8 пятерок, 13 четверок, 6 троек и 2 двойки. Вася получил тетрадь с оценкой?

(Ответ: 1,74 бита)

.

*

Вероятностный подход

Если $p_1 = p_2 = \cdots = p_N$, следовательно исходы равновероятны, то вероятность каждого исхода – это:

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

$$I = -\sum_{i=1}^{N} \frac{1}{N} \cdot \log_2 \frac{1}{N} = \log_2 N$$

$$2^I = N$$

формула **Хартли**

(Ральф Хартли американский инженер- связист) - предложена в 1928 г.

Задание для групп

1. По данному тексту составьте частотный словарь русского языка для букв, входящих в слово …….. Рассчитайте информационный объем этого слова по алфавитному и содержательному подходу к определению количества информации. Сделать выводы.
2. Сделайте выводы относительно зависимости информационных весов символов от их частотной характеристики и правила расположения букв на клавиатуре.

*

Результаты:

I группа:

Содержательный подход:
Всего символов в тексте : 2266

В	Е	Р	О	Я	Т	Н	С	Ь
61	172	95	163	33	130	176	103	15
0,027	0,076	0,042	0,072	0,015	0,057	0,077	0,045	0,006
5,22	3,72	4,58	3,80	6,10	4,13	3,69	1,14	7,24

Информационный объем=5,22+3,72+4,58+2*3,80+ 6,10+

+2*4,13+3,69+ 1,14+7,24 ≈ 47,5 бит

Алфавитный подход:
Мощность алфавита: N=33 I=5,04 бит

Информационный объем=5,04*11=55,44 бит

Результаты:

II группа:

Содержательный подход:
Всего символов в тексте : 2266

К	О	М	П	Ь	Ю	Т	Е	Р
81	163	84	30	15	14	130	172	95
0,036	0,072	0,37	0,013	0,007	0,006	0,057	0,076	0,042
4,81	3,80	4,75	6,24	7,24	7,34	4,13	3,72	4,58

Информационный объем=4,81+3,80+4,75+6,24+7,24+7,34+4,13+3,72+4,58=46,6 бит

Алфавитный подход:
Мощность алфавита: N=33 I=5,04 бит

Информационный объем=5,04*9=45,36 бит

*

Результаты:

III группа:

Содержательный подход:
Всего символов в тексте : 2266

И	Н	Ф	О	Р	М	А	Ц	Я
194	176	28	163	95	84	159	12	33
0,086	0,777	0,012	0,072	0,042	0,037	0,07	0,005	0,015
3,55	3,69	6,34	3,79	4,58	4,75	3,83	7,64	6,10

Информационный объем=2*3,55+3,69+6,34+3,79+4,58+4,375+ 3,83+7,64+6,10= 47,4 бит

Алфавитный подход:
Мощность алфавита: N=33 I=5,04 бит

Информационный объем=5,04*10=50,4 бит

*

Результаты:

Таблица частотности символов русского алфавита

	Символ	Р		Символ	Р		Символ	Р
1	о	0,070	12	л	0,035	23	ф	0,012
2	е	0,076	13	в	0,027	24	й	0,010
3	ё		14	д	0,025	25	х	0,009
4	а	0,070	15	п	0,016	26	ж	0,007
5	и	0,086	16	у	0,015	27	ь	0,007
6	н	0,077	17	я	0,014	28	ъ	0,006
7	т	0,057	18	б	0,014	29	ю	0,006
8	с	0,045	19	ы	0,013	30	ш	0.006
9	р	0,042	20	з	0.013	31	ц	0,004
10	м	0,037	21	г	0.012	32	щ	0,003
11	к	0,036	22	ч	0,012	33	э	0,003

*

Выводы:

- При алфавитном подходе, считается, что все символы алфавита встречаются в сообщениях, записанных с помощью этого алфавита, одинаково часто. Однако, в действительности символы алфавитов естественных языков в сообщениях появляются с разной частотой.

- Количество информации в сообщении о некотором событии зависит от вероятности этого события. Чем меньше вероятность некоторого события, тем больше информации содержит сообщение об этом событии.