Вероятностный подход к определению количества информации. Формула Шеннона

10 класс

Для вычисления количества информации в сообщении о неравновероятном событии используют следующую формулу:

$$I=\log_2(1/p)$$

где

I – это количество информации,

p – вероятность события.

Вероятность события выражается в долях единицы и вычисляется по формуле:

$$p=K/N$$

где

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

N – общее число возможных исходов какого-то процесса.

Задача: ***Бабушка испекла 8 пирожков с капустой, 24 пирожков с повидлом. Маша съела один пирожок. Вычислить вероятность выбора пирожка с разной начинкой и количество информации, которое при этом было получено.***

Решение:

Пусть K_1 – это количество пирожков с повидлом, $K_1=24$

K_2 – количество пирожков с капустой, $K_2=8$

N – общее количество пирожков, $N = K_1 + K_2 = 24+8=32$

Вероятность выбора пирожка с повидлом: $p_1=24/32=3/4=0,75$.

Вероятность выбора пирожка с капустой: $p_2=8/32=1/4=0,25$.

Обращаем внимание учащихся на то, что в сумме все вероятности дают 1.

Вычислим количество информации, содержащееся в сообщении, что Маша выбрала пирожок с повидлом:

$$I_1=\log_2(1/p_1)= \log_2(1/0,75)= \log_2 1,3=1,15470 \text{ бит.}$$

Вычислим количество информации, содержащееся в сообщении, если был выбран пирожок с капустой:

$$I_2=\log_2(1/p_2)= \log_2(1/0,25)= \log_2 4=2 \text{ бит.}$$

Качественную связь между вероятностью события и количеством информации в сообщении об этом событии можно выразить так: *чем меньше вероятность некоторого события, тем больше информации содержит сообщение об этом событии.*

Вопрос: сколько получим информации при выборе пирожка любого вида?

Ответить на этот вопрос нам поможет формула вычисления количества информации для событий с различными вероятностями, которую предложил в 1948 г. американский инженер и математик К.Шеннон.

Если **I**-количество информации, **N**-количество возможных событий, $\mathbf{p_i}$ - вероятности отдельных событий, где i принимает значения от 1 до N, то количество информации для событий с различными вероятностями можно определить

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

можно расписать формулу в таком виде:

$$\boxed{I = -(p_1 \cdot \log_2 p_1 + p_2 \cdot \log_2 p_2 + p_3 \cdot \log_2 p_3 + p_4 \cdot \log_2 p_4 + \ldots\ldots\ldots + p_N \cdot \log_2 p_N)}$$

Рассмотрим формулу на нашем примере:

$$I = -(p_1 \cdot \log_2 p_1 + p_2 \cdot \log_2 p_2)$$

$$= -(0{,}25 \cdot \log_2 0{,}25 + 0{,}75 \cdot \log_2 0{,}75)$$

$$\approx -(0{,}25 \cdot (-2) + 0{,}75 \cdot (-0{,}42))$$

$$= 0{,}815 \text{ бит}$$

Задача: В корзине лежат 32 клубка красной и черной шерсти. Среди них 4 клубка красной шерсти.

Сколько информации несет сообщение, что достали клубок красной шерсти? Сколько информации несет сообщение, что достали клубок шерсти любой окраски?

Применение ЭТ Excel для решения задач на нахождение количества информации

$$p_i = K_i / N$$

$$I_i = \log_2(1/p_i)$$

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

Задача №1

В озере обитает 12500 окуней, 25000 пескарей, а карасей и щук по 6250. Какое количество информации несет сообщение о ловле рыбы каждого вида. Сколько информации мы получим, когда поймаем какую-нибудь рыбу?

Дано: $K_о=12500$; $K_п=25000$; $K_к= K_щ=6250$

Найти: $I_о, I_п, I_к, I_щ, I$

Решение:

Найдем общее количество рыбы:

$$N= K_о+K_п+K_к+K_щ.$$

Найдем вероятность ловли каждого вида рыбы:

$$p_о= K_о/N;\ p_п= K_п/N;\ p_к= p_щ= K_к/N.$$

Найдем количество информации о ловле рыбы каждого вида:

$$I_о= \log_2(1/p_о);\ I_п=\log_2 (1/p_п);\ I_к= I_щ= \log_2 (1/p_к)$$

Найдем количество информации о ловле рыбы любого вида:

$$I= p_о\cdot\log_2 p_о+ p_п\cdot\log_2 p_п +p_к\cdot\log_2 p_к +p_щ\cdot\log_2 p_щ$$

При составлении таблицы мы должны учитывать:

- Ввод данных (что дано в условии).
- Подсчет общего количества числа возможных исходов **(формула $N=K_1+K_2+\ldots+K_i$).**
- Подсчет вероятности каждого события **(формула $p_i= K_i/N$).**
- Подсчет количества информации о каждом происходящем событии **(формула $I_i= \log_2(1/p_i)$).**
- Подсчет количества информации для событий с различными вероятностями **(формула Шеннона).**

	A	B	C	D	E	F
1	**События**	**Число событий**	**Вероятность (P=K/N)**	**Кол-во информ. о происходящих событиях $I=LOG_2(1/P)$**	**$P^*LOG_2(P)$**	**$I=-\sum P_i{}^*LOG_2P_i$**
2	окунь	12500	=B2/$B$6	=LOG(1/C2;2)	=C2*LOG(C2;2)	=-СУММ(E2:E5)
3	пескарь	25000	=B3/$B$6	=LOG(1/C3;2)	=C3*LOG(C3;2)	
4	карась	6250	=B4/$B$6	=LOG(1/C4;2)	=C4*LOG(C4;2)	
5	щука	6250	=B5/$B$6	=LOG(1/C5;2)	=C5*LOG(C5;2)	
6	Всего (N)	=СУММ(B2:B5)				

	A	B	C	D	E	F
1	**События**	**Число событий**	**Вероятность (P=K/N)**	**Кол-во информ. о происходящих событиях**	**$P^*LOG_2(P)$**	**$I=-\sum P_i{}^*LOG_2P_i$**
2	окунь	12500	0,2500	2,0000	-0,5000	1,7500
3	пескарь	25000	0,5000	1,0000	-0,5000	
4	карась	6250	0,1250	3,0000	-0,3750	
5	щука	6250	0,1250	3,0000	-0,3750	
6	Всего (N)	50000				

Практическая работа.

1.Сделать табличную модель для вычисления количества информации.

2. Используя табличную модель, сделать вычисления к задаче №2 результат вычисления занести в тетрадь.

Задача №2

В классе 30 человек. За контрольную работу по информатике получено 15 пятерок, 6 четверок, 8 троек и 1 двойка. Какое количество информации несет сообщение о том, что Андреев получил пятерку?

	A	B	C	D	E	F
1	**События**	**Число событий**	**Вероятность (P=K/N)**	**Кол-во информ. о происходящих событиях**	**$P*LOG_2(P)$**	**$I=-\sum P_i*LOG_2P_i$**
2	5	15	0,5000	1,0000	-0,5000	1,6365
3	4	6	0,2000	2,3219	-0,4644	
4	3	8	0,2667	1,9069	-0,5085	
5	2	1	0,0333	4,9069	-0,1636	
6	Всего (N)	30				

Домашняя работа

Задача№1

В коробке лежат кубики: 10 красных, 8 зеленых, 5 желтых, 12 синих. Вычислите вероятность доставания кубика каждого цвета и количество информации, которое при этом будет получено.

Задача№2

В непрозрачном мешочке хранятся 10 белых, 20 красных, 30 синих и 40 зеленых шариков. Какое количество информации будет содержать зрительное сообщение о цвете вынутого шарика?

shkola-ikt@bk.ru