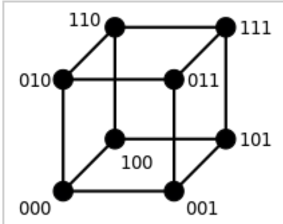
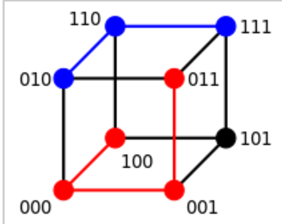
Расстояние Хэмминга

Материал из Википедии — свободной энциклопедии

Расстоя́ние Хэ́мминга — число позиций, в которых соответствующие символы двух слов одинаковой длины различны[1]. В более общем случае расстояние Хэмминга применяется для строк одинаковой длины любых q-ичных алфавитов и служит метрикой различия (функцией, определяющей расстояние в метрическом пространстве) объектов одинаковой размерности.



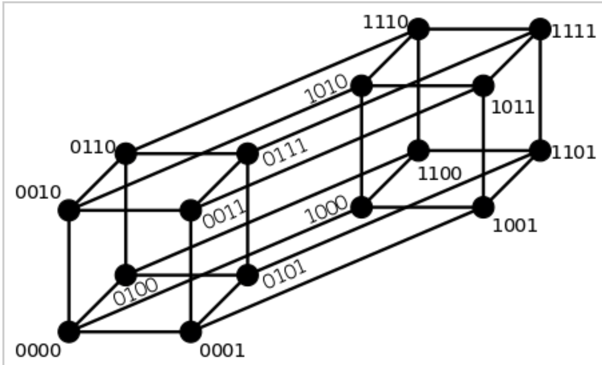
3-битовый двоичный куб. Все вершины, соединённые ребром имеют расстояние Хэмминга равное 1.



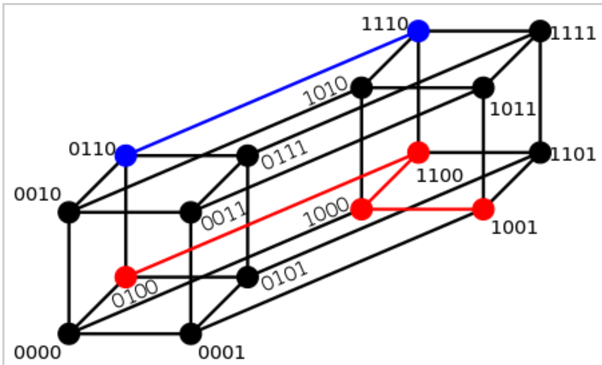
Пример двух расстояний: 100 → 011 имеют расстояние 3 (красные рёбра); 010 → 111 имеют расстояние 2 (синие рёбра).

Первоначально метрика была сформулирована Ричардом Хэммингом во время его работы в Bell Labs для определения меры различия между кодовыми комбинациями (двоичными векторами) в векторном пространстве кодовых последовательностей, в этом случае расстоянием Хэмминга $d(x, y)$ между двумя двоичными последовательностями (векторами) x и y длины n называется число позиций, в которых они различны — в такой формулировке расстояние Хэмминга вошло в словарь алгоритмов и структур данных национального института стандартов и технологий США (англ. *NIST Dictionary of Algorithms and Data Structures*). Расстояние Хэмминга является частным случаем метрики Минковского (при соответствующем определении вычитания):

$$d_{ij} = \sum_{k=1}^{p} |x_{ik} - x_{jk}|.$$



4-битовый двоичный четырёхмерный куб (тессеракт) где наглядны расстояния Хэмминга.



Примеры расстояний в двоичном тессеракте: 0100 → 1001 с расстоянием 3 (красные линии); 0110 → 1110 расстояние 1 (синяя линия).

В некоторых системах счисления, например, в коде Грея кодированные числа, различающиеся на 1, имеют расстояние Хемминга равное 1, говорят что такие числа являются «соседними», и вообще, два слова, не обязательно двоичные, расстояние Хемминга между которыми равно 1 называют «соседними».

Соседнее кодирование важно при проектировании логических устройств, где необходимо исключить логические гонки.

Примеры

- $d(1011101, 1001001) = 2$
- $d(2173896, 2233796) = 3$
- $d(toned, roses) = 3$

Свойства

Множество слов равной длины образуют метрическое пространство где для каждой пары элементов пространства определено число — расстояние Хэмминга $d(x, y)$ удовлетворяющее аксиомам метрики:

1. $d(x,\ y) = 0 \Leftrightarrow x = y$ (*аксиома тождества*).
2. $d(x,\ y) = d(y,\ x)$ (*аксиома симметрии*).
3. $d(x,\ z) \leqslant d(x,\ y) + d(y,\ z)$ (*аксиома треугольника* или неравенство треугольника).

- Из аксиом следует неотрицательность расстояния, поскольку

  $$0 = d(x,\ x) \leqslant d(x,\ y) + d(y,\ x) = 2{\cdot}d(x,\ y)\,.$$

- Если неравенство треугольника представить в виде

  $d(x,y) \leqslant d(x,z) + d(y,z)$ для всех x, y и z,

  тогда из аксиомы тождества и неравенства треугольника следует аксиома симметрии.

Расстояние Хэмминга всегда:

$$d(x,\ y) \leqslant n,$$

где n — длина слов в символах.

Расстояние Хэмминга в биоинформатике и геномике

Для нуклеиновых кислот (ДНК и РНК) возможность гибридизации двух полинуклеотидных цепей с образованием вторичной структуры — двойной спирали — зависит от степени комплементарности нуклеотидных последовательностей обеих цепей. При увеличении расстояния Хэмминга количество водородных связей, образованных комплементарными парами оснований уменьшается и, соответственно, уменьшается стабильность двойной цепи. Начиная с некоторого граничного расстояния Хэмминга гибридизация становится невозможной.

При эволюционном расхождении гомологичных ДНК-последовательностей расстояние Хэмминга является мерой, по которой можно судить о времени, прошедшем с момента расхождения гомологов, например, о длительности эволюционного отрезка, разделяющего гены-гомологи и ген-предшественник.

См. также

- Расстояние Левенштейна
- Обнаружение и исправление ошибок
- Матрица расстояний

Примечания

1. ↑ *Hamming distance: The number of digit positions in which the corresponding digits of two binary words of the same length are different (Federal Standard 1037C).*

Литература

- *Блейхут Р.* Теория и практика кодов, контролирующих ошибки = Theory and Practice of Error Control Codes. — М.: Мир, 1986. — 576 с.