Theory: Hamming distance

4 minutes 6 / 6 problems solved

Start practicing

§1. Formulating the problem

Sometimes to solve a problem one needs to identify whether two strings are similar. For example, assume we want to compare two DNA sequences to understand how different two organisms are. Or we want to find a word in a document, but not only exact but approximate matchings as well. To solve these problems, we need to come up with a metric for identifying the similarity between two strings and implement an algorithm for calculating the metric.

For now, there exist several metrics that allows estimating the similarity for two strings. In this lesson, we will consider one of the simplest metrics called the **Hamming distance**. For two strings of equal length, the Hamming distance is the number of positions in which the strings are different. Or, in other words, the metric measures the minimal number of substitutions required to transform one string into the other. It is used in information theory to measure how similar two binary codes are, in telecommunication to count the number of different bits in two binary words as an estimate of an error (some other examples may be found in a Wikipedia article).

§2. Algorithm description

An algorithm for calculating the Hamming distance is pretty straightforward: let's compare corresponding symbols of two strings from the first to the last symbol and count the number of different ones. The total number of different symbols is the Hamming distance for the strings.

§3. An example

Given below is an example of calculation of the Hamming distance for strings "microscope" and "microphone". Equal symbols are shown in green, different ones are shown in red.

M	I	C	R	O	S	C	O	P	E
M	I	C	R	O	P	H	O	N	E

As you can see, there are three different symbols. So, the Hamming distance is equal to 3.

§4. Complexity analysis

The running time of the algorithm is $O(n)$ where n is the number of symbols in one of the strings. The algorithm requires no additional memory.

Report a typo

43 users liked this theory. 1 didn't like it. What about you?

Start practicing

Comments (0) Hints (0) Useful links (0) Show discussion

Current topic:

Hamming distance

Topic depends on:

String basics

Topic is required for:

Hamming distance in Java

Table of contents: