Abstract:

The Hamming and the edit metrics are two common notions of measuring distances between pairs of strings \( x, y \) lying in the Boolean hypercube. The edit distance between \( x \) and \( y \) is defined as the minimum number of character insertion, deletion, and bit flips needed for converting \( x \) into \( y \). Whereas, the Hamming distance between \( x \) and \( y \) is the number of bit flips needed for converting \( x \) to \( y \).

In this paper we study a randomized injective embedding of the edit distance into the Hamming distance with a small distortion. This question was studied by Jowhari (ESA 2012) and is mainly motivated by two questions in communication complexity: the document exchange problem and deciding edit distance using a sketching protocol.

We show a randomized embedding with quadratic distortion. Namely, for any \( x,y \) satisfying that their edit distance equals \( k \), the Hamming distance between the embedding of \( x \) and \( y \) is \( O(k^2) \) with high probability. This improves over the distortion ratio of \( O(\log n \log^* n) \) obtained by Jowhari for small values of \( k \). Moreover, the embedding output size is linear in the input size and the embedding can be computed using a single pass over the input.