Properties of Near Optimal Edit Metric Error Correcting Codes

Lachlan Plant*, Sheridan Houghten, Brock University

Detection and correction of errors within communications using error correcting codes allows for safer transmission over noisy channels. Generation of these codes however can be extremely time consuming with more complex types of errors such as insertion and deletion of bits. This research looks at the codes with the maximal number of codewords, based on the Levenstein distance metric between the words (Edit codes), in order to find properties which can be exploited in generation. It has been conjectured that there exists a balanced Hamming code of maximal size for every parameter set. This research disproves the conjecture when this idea is applied to edit codes. Codes of maximal size were completely generated up to length 9 and distance 3, at which point enumeration becomes intractable, and analysed for the balance of the code, as well as the codes prefix balance. Codes are considered balanced if in each position of the code, the total number of each symbol in the codewords are equal, and are considered prefix balanced if the total number of codewords with prefix $x_0...x_{n/2}$ is equal to those with the inverse prefix. After full generation no balanced maximal $(9, 3)_2$ codes exist, disproving the idea that there will exist a maximal balanced code for all parameter sets. For all other parameter sets, and near maximal $(9, 3)_2$ codes, balanced codes and prefix balanced codes were found to exist. The reduced search space of the prefix balance codes was then used to search for and find a $(10, 3)_2$ code of optimally known size.

Keywords: Error Correcting Code, Levenstein Distance, Code Balance, Edit Codes