# Title

Bag of Tricks for Efficient Text Classification

# Author

Armand Joulin, Edouard Grave, Piotr Bojanowski, Tomas Mikolov

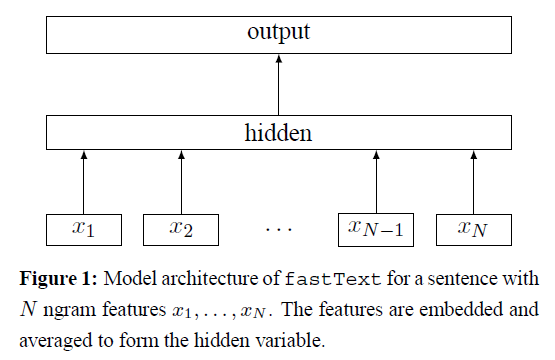
# Abstract

Authors propose a fast text classifier, **fastText**, which is **on par** with deep learning classifiers in terms of accuracy and **many orders of magnitude faster** for training and evaluation.

# Issue

1. Models based on neural networks tend to be relatively **slow** both at train and test time, limiting their use on very large datasets.
2. Linear classifiers need **right features** to obtain state-of-the-art performances.

# Method



1. Using a look-up table, **bags of ngram** covert to **word representations**.
2. Word representations are **averaged** into a text representation, which is an hidden variable.
3. Text representation is in turn fed to a linear classifier.
4. Use the **softmax** function f to compute the probability distribution over the predefined classes.

For a set of documents, this leads to minimizing the **negative log-likelihood** over the classes:

where is the normalized bag of features of the nth document, the label, and the weight matrices.

## Hierarchical softmax

In order to improve our running time, authors use a **hierarchical softmax** based on the Huffman coding tree.

## N-gram features

Authors use a **bag of n-grams** as additional features to capture some partial information about the local word order.