Easy Data Augmentation (EDA) for NLP

- EDA is a set of simple and universal data augmentation techniques for NLP tasks, specifically text classification.
- EDA includes four text editing operations: synonym replacement, random insertion, random deletion, and random swap.
- EDA can help to overcome the problem of limited training data, particularly for smaller datasets.

EDA Evaluation

- EDA was evaluated on five benchmark classification tasks, showing substantial improvements on all five tasks.
- EDA outperformed other data augmentation techniques, such as back-translation and word embedding-based methods, in terms of accuracy.
- EDA was tested on both recurrent neural networks and convolutional neural networks.

EDA Implementation

- EDA is easy to implement and can be applied to existing text data to generate new samples.
- EDA code is publicly available on GitHub.
- EDA can be used to improve the performance of text classification models, particularly for smaller datasets.

EDA Operations

- Synonym replacement: replaces words in the text with their synonyms.
- Random insertion: inserts random words into the text.
- Random deletion: deletes random words from the text.

EDA Operations (cont.)

- Random swap: swaps two words in the text.
- EDA operations are simple and can be easily applied to existing text data.
- EDA operations can generate new samples, which can help to overcome the problem of limited training data.