**1. Tell me why I am wrong? Explaining errors to text classifiers**

Text classifiers help us to extract useful information from text by identifying relevant documents and categorising them. For example, in a crisis situation, a classifier can find reports of emergencies on social media and forward them to the relevant agencies [1, 2]. However, text classifiers need large amounts of labelled training data to achieve high accuracy, which are often unavailable. This project investigates an alternative: a user corrects the classifier’s errors by explaining how they identified the correct label. An approach for learning from explanations was previously shown to improve classifier performance [3], but there are several ways this approach could be improved. This project aims to test some possible improvements and use the idea to perform human-in-the-loop error correction.

1. TREC Incident Streams: Finding Actionable Information on Social Media: <https://eprints.gla.ac.uk/183409/>

Affect elements: data quality(standardizing); insufficient trained staff; policies;

how well did they identify the information type of the tweet (Information Type Categorization); and how accurately did they estimate the criticality of the information within each tweet (Information Criticality)

1. CrisisNLP: <https://crisisnlp.qcri.org/humaid_dataset>

This is a dataset.

tweet\_id tweet\_text class\_label

733360620339257344 .@GreenABEnergy How can @AirworksCanada assist in the cleanup? #AlbertaStrong rescue\_volunteering\_or\_donation\_effort

Humanitarian categories

\* Caution and advice

\* Displaced people and evacuations

\* Dont know cant judge

\* Infrastructure and utility damage

\* Injured or dead people

\* Missing or found people

\* Not humanitarian

\* Other relevant information

\* Requests or urgent needs

\* Rescue volunteering or donation effort

\* Sympathy and support

1. ExpBERT: <https://arxiv.org/abs/2005.01932>

ExpBERT: ExpBERT: Representation Engineering with Natural Language Explanations which use less sample and with high f1 score.

BERT: include [CLS] [SEP] and [PAD],where [CLS] is on the start of the sentence and [SEP] is used to split the sentences and [PAD] is used to cut the sentence into same length.

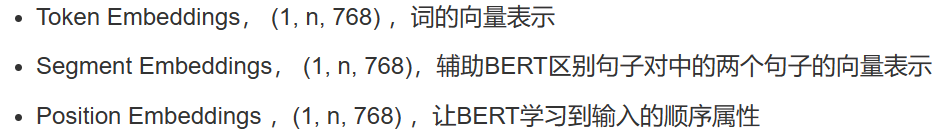
BERT: use token embedding segment embedding and position embedding

Use nature language *explanations generate features* can improve the modern neural representations and restore the control to increase the expressive power.

{s,o1,o2} s is s sequence of words and o1 and o2 are two entities (o1:Ben,o2:Lindy)

Title: Improvement of nature language explanations

<https://github.com/MurtyShikhar/ExpBERT>



但是从从特征工程向端到端神经网络和表示学习的转变减轻了手动特征工程的负担并提高了模型的表达能力，但它也减少了人们对模型归纳偏差的控制。

## Active Learning Literature Survey

https://minds.wisconsin.edu/handle/1793/60660