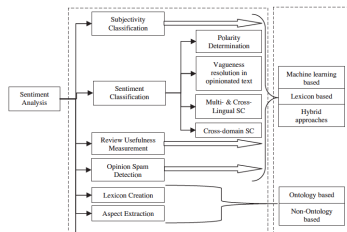# Sentimental Analysis with CharSCNN - Introduction

Sentiment analysis is widely used in e-commerce, political movements, company strategies, marketing campaigns, algorithm trading.

- Sentimental Analysis is challenging due to unstructured and noisy data
- CharSCNN is used in polarity determination in this project. The model structure use convolutional layers and recurrent neural networks to exploit from character to sentence level representation and reach comparable results.

Sentimental Analysis with CharSCNN - Methodology

Dataset:

Stanford Twitter Sentiment corpus(STS) contains 1.6 million tweets and labelled as positive/negative. In our experiment, we sampled 160k tweets as the dataset we use.

Finance bank phrase includes 2264 news titles with sentiment labels as positive, neutral and negative.

Dataset	#sentence	#Classes
STS	16000	2
Financial PhraseBank	2264	3

Sentimental Analysis with CharSCNN - Methodology

Model:

- **Word Embeddings** use an encoded matrix $w^{wrd} \in \mathcal{R}^{d^{wrd}*V^{wrd}}$ to represent single words in the vocabulary.
- **Character embeddings** use a numerical vector $r^{chr} \in d^{chr}$ to represent characters. Here we apply a Long Short Term Memory Netork(LSTM) on the character embedding vectors $\{r_1^{chr}, r_2^{chr}, ..., r_M^{chr}\}$ for a given word to produce a local character feature r^{wch}.
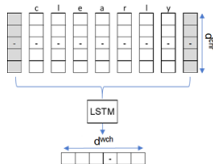

- **Sentence-level Embeddings** combines word and character-level embeddings and the joint vectors are $\{u_1, u_2, ..., u_N\}$ for a given sentence x with N words. Here $u_n \in \mathcal{R}^{(d^{wrd}+d^{wch})}$. Then a convolutional layer is used to produce the sentence representation.
- Dense layers for classification: $s(x) = W^3 h(W^2 r_x^{sent} + b^2) + b^3$

Sentimental Analysis with CharSCNN - Result

The CharSCNN works well on both data.

- The test accuracy on STS data is around 73%.
- The test accuracy on Financial Phrase Bank is around 70%.
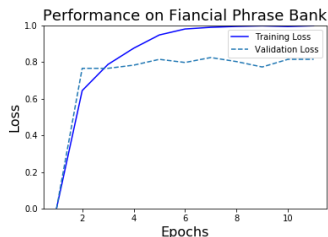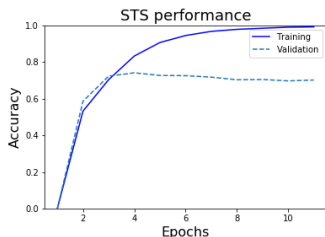
Sentimental Analysis with CharSCNN - Discussion

- The results on both datasets show that CharSCNN is an effective model in sentimental classification.
- The CharSCNN can capture the pattern from character-level to a sentence-level using convolutional neural networks, which is effective to present the semantics regarding sentiment.
- Future work:
  - Analyze more details of the character level representation combining the linguisitcs domain knowledge.
  - Apply the model on financial news data such as Reuters to compare with the trend with the stock market.