

# Lancaster A at SemEval-2017 Task 5: Evaluation metrics matter: predicting sentiment from financial news headlines

Andrew Moore and Paul Rayson

School of Computing and Communications, Lancaster University, UK

@apmoore94 @perayson

## Pre-Processing

- Lower cased.
- Tokenized using Unitok.

## Problem

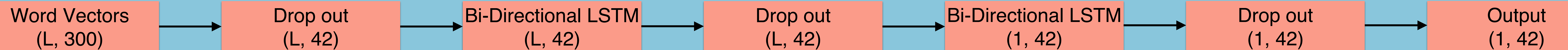
Predict the sentiment of financial headlines with respect to a company mentioned within the headlines.

## Data

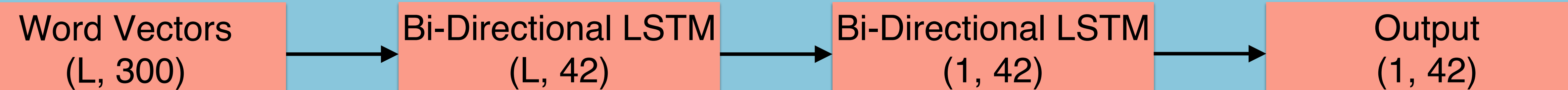
- Financial Word Embeddings<sup>1</sup>.
- 1142 Training examples.
- 491 test examples

## Methods

### Early Stopping Bi-Direction LSTM (ELSTM)



### Standard Bi-Direction LSTM (SLSTM)



### SVR. With the following features:

- Uni- and Bi- grams word representations.
- Special word representations where company names, positive and negative words were replaced with special tokens respectively.
- The aspect (company name) of the sentence of which the sentiment is with respect to.

## Evaluation

Metric 1

Cosine Similarity (CS)

$$\frac{\sum_{i=1}^K y_i \hat{y}_i}{\sqrt{\sum_{i=1}^K y_i^2} \sqrt{\sum_{i=1}^K \hat{y}_i^2}}$$

Metric 2

$$\frac{\sum_{n=1}^N CS(\hat{y}_n, y_n)}{N}$$

**K = Total number of samples, N = Total number of sentences**

Metric 3

$$\frac{\sum_{n=1}^N \begin{cases} len(\hat{y}_n) * CS(\hat{y}_n, y_n), & \text{if } len(\hat{y}_n) > 1 \\ 1 - |y - \hat{y}_n|, & \text{if } \frac{\hat{y}_n}{y} \geq 0 \end{cases}}{K}$$

## Results

	Metric		
Model	1	2	3
SVR	62.14	54.59	62.34
SLSTM	72.89	61.55	68.64
ELSTM	<b>73.20</b>	<b>61.98</b>	<b>69.24</b>

1. [https://github.com/apmoore1/semeval/tree/master/models/word2vec\\_models](https://github.com/apmoore1/semeval/tree/master/models/word2vec_models)

Code available here: <https://github.com/apmoore1/semeval>