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

Andrew Moore and Paul Rayson
February 27, 2018
School of Computing and Communications, Lancaster University.
Task
The task

Example sentence

‘Why AstraZeneca plc & Dixons Carphone PLC Are Red-Hot Growth Stars!’

Sentiment scale

Data
Training data: 1142 samples, 960 headlines/sentences.
Testing data: 491 samples, 461 headlines/sentences.
Approach
Models

1. Support Vector Regression (SVR) [1]
2. Bi-directional Long Short-Term Memory BLSTM [2][3]
Pre-Processing and Additional data used

Pre-Processing

1. Lower cased.
2. Tokenised.

Word2Vec model

Used 189, 206 financial articles (e.g. Financial Times) that were manually downloaded from Factiva\(^1\) to create a Word2Vec model [5]\(^2\).

These were created using Gensim\(^3\).

\(^1\)https://global.factiva.com/factivalogin/login.asp?productname=global
\(^2\)https://github.com/apmoore1/semeval/tree/master/models/word2vec_models
\(^3\)https://radimrehurek.com/gensim/models/word2vec.html
Features and settings that we changed

1. Tokenisation - Whitespace or Unitok\(^4\)
2. N-grams - uni-grams, bi-grams and both.
3. SVR settings - penalty parameter C and epsilon parameter.
4. Target aspect.
5. Word Replacements.

\(^4\text{http://corpus.tools/wiki/Unitok}\)
Example Sentence

‘AstraZeneca PLC had an improved performance where as Dixons performed poorly’

‘companyname had an posword performance where as companyname performed negword’
Two BLSTM models

**Standard Model (SLSTM)**
- Drop out between layers and connections.
- 25 times trained over the data (epoch of 25).

**Early stopping model (ELSTM)**
- Drop out between layers only.
- Early stopping used to determine the epoch.
Loss function
Mean Square Error (MSE)

$$\frac{1}{Y} \sum_{i=1}^{Y} (\hat{y}_i - y)^2$$ (1)
Findings and Results
SVR best features

Features

- Using uni-grams and bi-grams to be the best. 2.4% improvement over uni-grams.
- Using a tokeniser always better. Affects bi-gram results the most. 1% improvement using Unitok\(^5\) over whitespace.
- SVR parameter settings important 8% difference between using $C=0.1$ and $C=0.01$.
- Incorporating the target aspect increased performance. 0.3% improvement.
- Using all word replacements. $N=10$ for POS and NEG words and $N=0$ for company. 0.8% improvement using company and 0.2% for POS and NEG.

\(^5\)http://corpus.tools/wiki/Unitok
The three different metrics

Cosine Similarity (CS) Metric 2

Metric 1

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

(2) Metric 3

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

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

(3)

\[
\frac{1}{K}
\]

(4)

\(K = \text{Total number of samples.}\)

\(N = \text{Total number of sentences.}\)
## Results across the different metrics

<table>
<thead>
<tr>
<th>Model</th>
<th>Metric 1</th>
<th>Metric 2</th>
<th>Metric 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>62.14</td>
<td>54.59</td>
<td>62.34</td>
</tr>
<tr>
<td>SLSTM</td>
<td>72.89</td>
<td>61.55</td>
<td>68.64</td>
</tr>
<tr>
<td>ELSTM</td>
<td>73.20</td>
<td>61.98</td>
<td>69.24</td>
</tr>
<tr>
<td>Fortia-FBK[4]</td>
<td>74.50</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Metric 1 was the final metric used.
‘uk stocks little changed as ashtead gains, housing shares drop’
Predicted: -0.43, Real: 0.23

‘standard life chief agrees 600000 bonus cut’
Predicted: -0.54, Real: 0.08

‘why i would put j sainsbury plc in my trolley before wm morrison supermarkets ...’
Predicted: 0.11, Real: 0.76
Future Work
1. Incorporate aspects into the BLSTM's shown to be useful by Wang et al. [7].
2. Improve BLSTM’s by using an attention model Wang et al. [7].
3. Add known financial sentiment lexicon into the LSTM model [6].
1. BLSTM outperform SVRs with minimal feature engineering.
2. The future is to incorporate more financial information into the LSTM’s.
Questions?

a.moore@lancaster.ac.uk       @apmoore94
p.rayson@lancaster.ac.uk      @perayson

All the code can be found here\(^6\)

Presentation can be found here \(^7\)

\(^6\)https://github.com/apmoore1/semeval
\(^7\)https://github.com/apmoore1/semeval/blob/master/presentation/semeval.pdf
**Support vector regression machines.**  

A. Graves and J. Schmidhuber.  
**Framewise phoneme classification with bidirectional Lstm and other neural network architectures.**  
*Neural Networks, 18(5):602–610, 2005.*

S. Hochreiter and J. Schmidhuber.  
**Long short-term memory.**  


Y. Wang, M. Huang, x. zhu, and L. Zhao.  
**Attention-based lstm for aspect-level sentiment classification.**  