Assessing Humour in Edited News Headlines using Hand-Crafted Features and Online Knowledge Bases

Bachelor Thesis 2020

Kristian Nørgaard Jensen, Marco Placenti, Nicolaj Filrup Rasmussen, Thai Wang

June 9, 2020

IT University of Copenhagen

Outline of Presentation

Table of Contents

(I) Humour Regression

Introduction

(I) Humour Regression

(II) Score $0 - 3$

(I) Humour Regression
(II) Score 0 – 3
(III) SemEval - 2020 [2]

Table of Contents

Data set

Headlines	Score
Sam's Club closes hundreds of ~~stores~~ carts nationwide.	0.2
Goldman ~~warns~~ dances on irreversible *Brexit* plans	0.4
Uber CTO blasts ~~Trump~~ techno in staff email	1.4
Elon Musk has just blasted the world's most powerful rocket into ~~space~~ wall.	2.4
Recent Scandals Highlight *Trump*'s Chaotic ~~Management~~ Fashion Style	3.0

Data set produced by [1]

Table of Contents

Research Question 1. Is it possible to use hand crafted features to aid in a humour recognition task?

Research Questions

Research Question 1. Is it possible to use hand crafted features to aid in a humour recognition task?

Hypothesis 1.1. By encoding meta information like sentence length, it is possible to inject useful information into a humour recognition model.

Research Question 1. Is it possible to use hand crafted features to aid in a humour recognition task?

Hypothesis 1.1. By encoding meta information like sentence length, it is possible to inject useful information into a humour recognition model.

Hypothesis 1.2. Using hand-crafted features can help in capturing structures of varying kinds of humour, thus allowing the model to generalise better.

Research Questions

Research Question 2. Is it possible to enhance humour recognition systems by using Knowledge bases in order to encode domain knowledge?

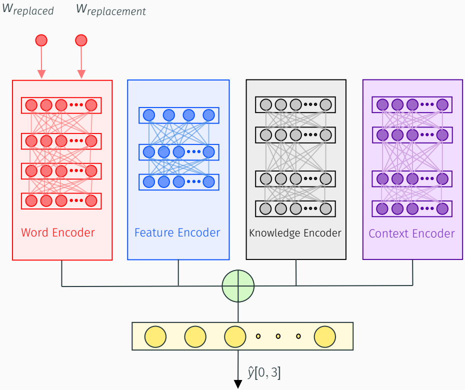
Research Questions

Research Question 2. Is it possible to enhance humour recognition systems by using Knowledge bases in order to encode domain knowledge?

Hypothesis 2.1. Seeing as we are using the data produced by [1], which stems from 2017-18, background knowledge is necessary in order to properly model the humour found in the data set.

Table of Contents

Model



$w_{replaced}$
$w_{replacement}$
Word Encoder
Feature Encoder
Knowledge Encoder
Context Encoder
$\hat{y}[0,3]$

Table of Contents

Phonetic Distance

Replaced/replacement	Levenshtein distance
‘Syria’ → `` `S IH1 R IY0 AH0' ``	0.1176
‘cereal’ → `` `S IH1 R IY0 AH0 L' ``	
‘coup’ → `'K UW1'`	0.9474
‘ignorance’ → `` `IH1 G N ER0 AH0 N S' ``	

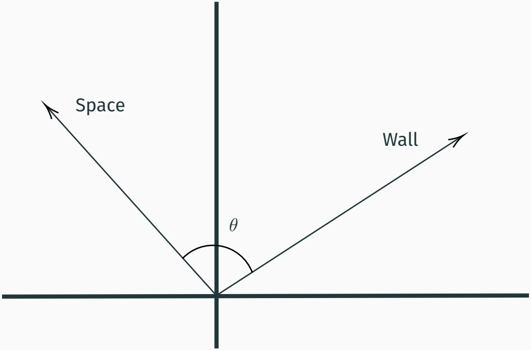
Relative position

Headline examples	Value
Sam's Club closes hundreds of ~~stores~~ carts nationwide.	0.86
Goldman ~~warns~~ dances on irreversible Brexit plans	0.33
Uber CTO blasts ~~Trump~~ techno in staff email	0.57
Elon Musk has just blasted the world's most powerful rocket into ~~space~~ wall.	1.00
Recent Scandals Highlight Trump's Chaotic ~~Management~~ Fashion Style	0.86

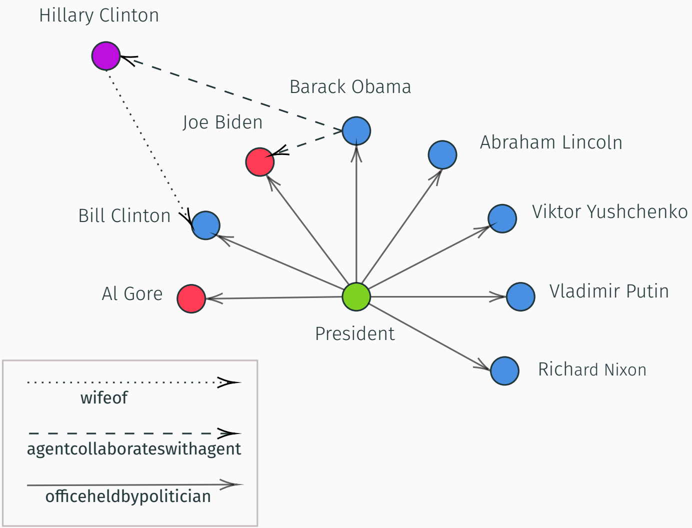
Sentence Length

Headline examples	Value
Sam's Club closes hundreds of ~~stores~~ carts nationwide.	0.35
Goldman ~~warns~~ dances on irreversible Brexit plans	0.30
Uber CTO blasts ~~Trump~~ techno in staff email	0.35
Elon Musk has just blasted the world's most powerful rocket into ~~space~~ wall.	0.60
Recent Scandals Highlight Trump's Chaotic ~~Management~~ Fashion Style	0.35

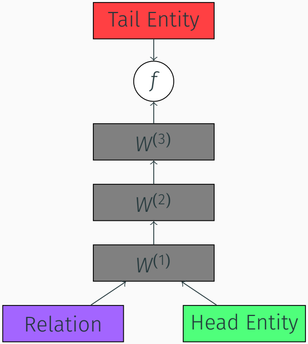
Cosine Similarity



Knowledge Vectors



Neural Association Model [3]



	Min	Average	Median	Max
NELL	0	4.8	5	13
NELL + WordNet	1	7.2	7	18

Evaluation

Root Mean Squared Error (RMSE)

$$\sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

Results

System	Test Score	Stddev
Mean Baseline	0.57471	(N/A)
Median Baseline	0.59140	(N/A)
Linear Regression	0.57361	(N/A)
CNN W. Highway Network	0.57543	0.00064
Multi-Channel CNN	0.59580	0.00510
KBLSTM	0.57304	0.00119
NNLM	0.56211	0.00111
Base model	0.55510	0.00168
Base model rounded[1]	0.55440	0.00200
Base model Albert Word Encoder[2]	0.57397	0.00066
HP Tuned base model	0.54476	0.00214
Extended model	0.54576	0.00145
HP Tuned extended model	0.54441	0.00214

Table of Contents

Summary

Key-points

Key-points

(I) Hand-Crafted Features

Key-points

(I) Hand-Crafted Features

- Some decent results

Key-points

(I) Hand-Crafted Features

- Some decent results
- Diving in closer might reveal better features

Summary

Key-points

(I) Hand-Crafted Features

- Some decent results
- Diving in closer might reveal better features

(II) Knowledge Encoder

Summary

Key-points

(I) Hand-Crafted Features

- Some decent results
- Diving in closer might reveal better features

(II) Knowledge Encoder

- Encodes the right information

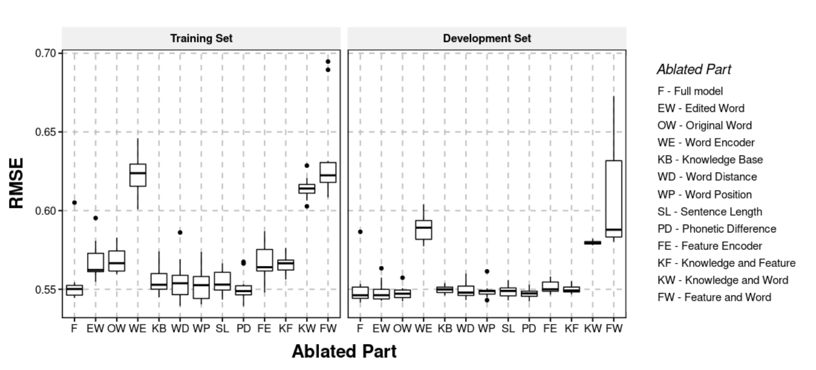
Key-points

(I) Hand-Crafted Features
- Some decent results
- Diving in closer might reveal better features

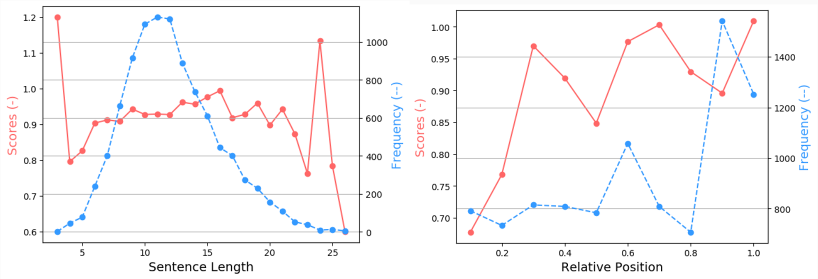
(II) Knowledge Encoder
- Encodes the right information
- Not incorporated correctly

Thank You!

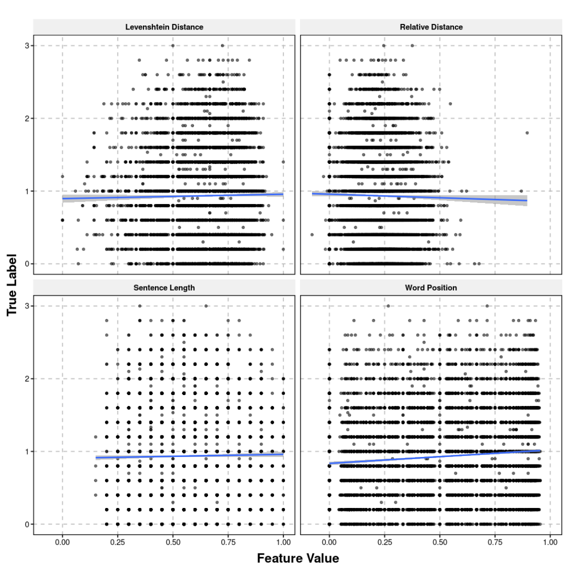
Ablation Study



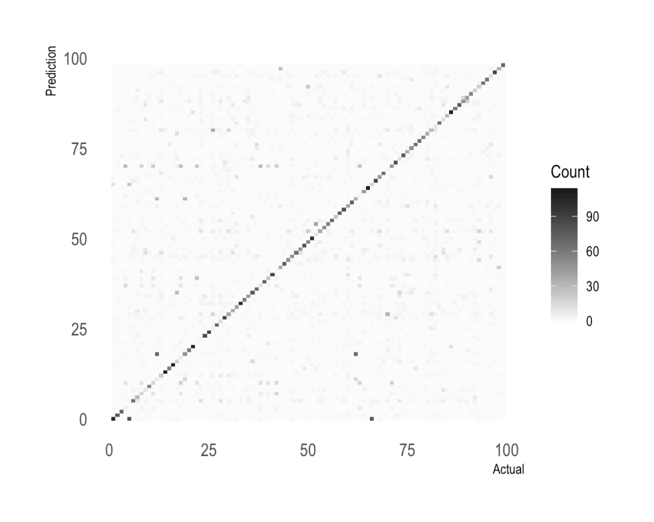
Sentence Length & Relative Position



Features


Levenshtein Distance
Relative Distance
Sentence Length
Word Position
True Label
Feature Value
0
1
2
3
0.00
0.25
0.50
0.75
1.00

Probing Encoded Knowledge



N. Hossain, J. Krumm, and M. Gamon.
“president vows to cut <taxes> hair”: Dataset and analysis of creative text editing for humorous headlines.
In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 133–142, Minneapolis, Minnesota, June 2019.

N. Hossain, J. Krumm, M. Gamon, and H. Kautz.
Semeval-2020 Task 7: Assessing humor in edited news headlines.
In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2020)*, Barcelona, Spain, 2020.

Q. Liu, H. Jiang, A. Evdokimov, Z.-H. Ling, X. Zhu, S. Wei, and Y. Hu.
Probabilistic reasoning via deep learning: Neural association models.
arXiv preprint arXiv:1603.07704, 2016.