
Pun Classification and Location

Shantanu Karnwal shantanu.karnwal@colorado.edu

Amir Kashipazha amirhossein.kashipazha@colorado.edu

Brandon Boylan-Peck brbo9266@colorado.edu

Cathlyn Stone
cathlyn.stone@colorado.edu

Kenneth Hunter Wapman
kennethwapman@colorado.edu

Abstract

Pun identification and location are challenging natural language processing tasks. We implemented several algorithms for both, with results which were comparable to those outlined in the SemEval conference which originally defined the tasks.

1 Overview

1.1 Pun Structure

here we will talk about what puns are

- homographic
- heterographic
- other types

here we will talk about the problems we worked on.

homographic pun example

heterographic pun example

Please read carefully the instructions below and follow them faithfully.

1.2 Motivation

1.3 Tasks

section references 2, 3, and 5 below.

2 Pun Detection

here we will talk about pun detection and our algorithms.

2.1 Baseline

2.2 Algorithms

2.3 Feature Comparison

3 Pun Location

3.1 Baseline

3.2 Algorithms

3.3 Feature Comparison

We introduced various features to improve accuracy. Table (number ?) shows the them and their indexes. These indexes are represent their related feature in feature related figures. Features were added to the code one by one to find how they improved accuracy. Accuracy change behave differently against each feature. For example, accuracy did not increased or decreased with positive and negative. In the other hand, unigram hiked up accuracy to 100 percent which means over fitting. We eliminated features which did effect accuracy from our code. Finally, we ended up using 11 features, index 0 to 10, for feature comparison.

In order to find what combinations of feature makes the largest accuracy, we used combination rule to find all possible combination of our 11 features. This made 2047 feature combination. We run code for each combination separately for homographic and heterographic pun, 4089 for both of them. Figures (add figure numbers here) shows the 10 largest accuracy of testing and training set for homographic and heterographic.

Index	Feaute Name
0	Lesk Algorithm
1	Pos
2	tfidf
3	Embeddings
4	Unigram
5	Number of Homophone in Pun
6	Number of Each Homophone in Pun
7	Homophone is in Pun or Not
8	Idiom is in Pun or Not
9	Antonyms is in Pun or Not
10	Homonym is in Pun or Not
11	Bigram
12	Trigram
13	Positives
14	Negatives
15	All or First Caps

4 Results

4.1 Pun Detection

Here's how we think we did.

4.1.1 Evaluation

4.1.2 Results

here are our results. here's the baseline. here's semeval's

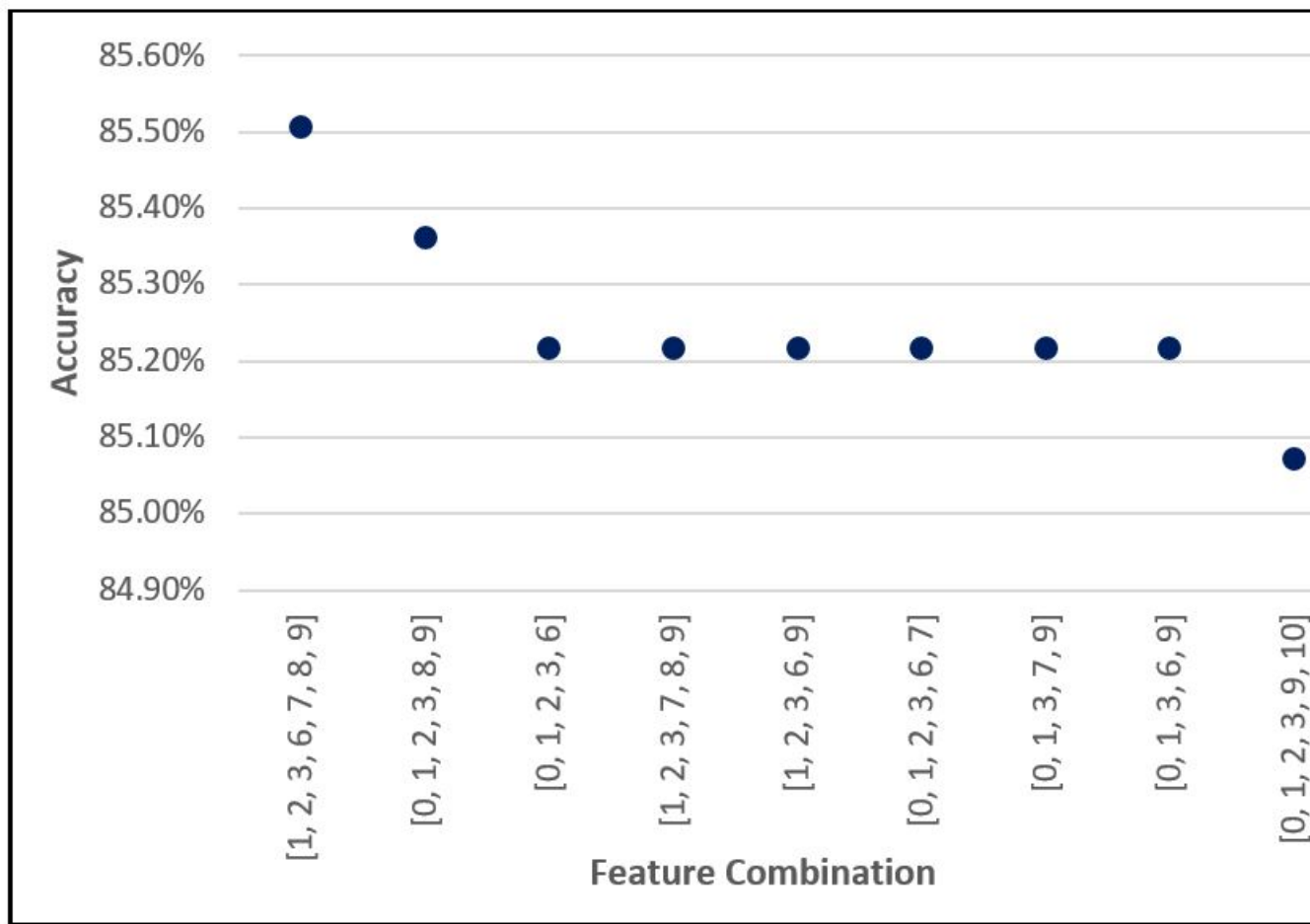


Figure 1: Figure NUMBER ? - Accuracy on Test Set for Homographic Pun

4.1.3 Error Analysis

4.2 Pun Location

Here's how we think we did.

4.2.1 Evaluation

f score etc etc

4.2.2 Results

here are our results. here's the baseline. here's semeval's

4.2.3 Error Analysis

5 Conclusion

5.1 who did what

5.2 what went well

5.3 what we could have done better

Acknowledgments

here's where we acknowledge stuff

References

- [1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.
- [2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural Simulation System*. New York: TELOS/Springer-Verlag.
- [3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.

6 Style Stuff

this is code-ish looking

here's a percent: $\sim 15\%$ **boldboldbold** *italics*

here's a fraction: $\frac{1}{4}$

This is paragraphed

`\citet{hasselmo}` investigated\dots

produces

Hasselmo, et al. (1995) investigated...

ref number: [4]

footnotes.¹

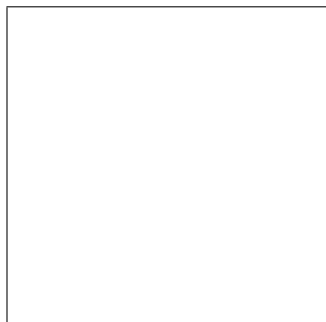


Figure 2: Sample figure caption.

Table 1.

use booktabs package for tables

¹they go after the period

Table 1: Sample table title

Part		
Name	Description	Size (μm)
Dendrite	Input terminal	~ 100
Axon	Output terminal	~ 10
Soma	Cell body	up to 10^6

<https://www.ctan.org/pkg/booktabs>