

# Stochastic Gradient Descentimiento Analysis

*Classifying Affect in Mixed Language Tweets*

*Connor Boyle  
Martin Horst  
Nikitas Tampakis*

*LING 573  
April 27, 2021*

# Roadmap

Intro to SemEval 2020

Dataset examples

System Architecture

Results

Discussion

Future steps

# **SemEval-2020 Task 9: Overview of Sentiment Analysis of Code-Mixed Tweets**

**Parth Patwa<sup>1\*</sup>   Gustavo Aguilar<sup>2\*</sup>   Sudipta Kar<sup>2</sup>   Suraj Pandey<sup>3</sup>  
Srinivas PYKL<sup>1</sup>   Björn Gambäck<sup>4</sup>   Tanmoy Chakraborty<sup>3</sup>  
Thamar Solorio<sup>2</sup>   Amitava Das<sup>5</sup>**

- Sentiment analysis on (Spanglish, Hinglish) tweets
- Categorical output: [-, 0, +]
- Evaluated on F1 score (weighted by category)

# Example Instance 1

Input: tokenized, language-tagged tweets, e.g.:

```
[  
  ("freshman", ne),  
  ("15", other),  
  ("empiezan", lang2),  
  ("hoy", lang2),  
  ("??", other),  
  ("holy", lang1),  
  ("yезus", mixed),  
  ("the", lang1),  
  ("food", lang1),  
  ("here: ')", lang1)  
]
```

Output: sentiment label

"positive"

## Example Instance 2

Input: tokenized, language-tagged tweets, e.g.:

```
[
  ("@_itskayyyyy", other),
  ("no", lang2),
  ("pos", lang2),
  ("wow", lang1),
  ("", other),
  ("Fuhk", lang1),
  ("the", lang1),
  ("@dolphins", other),
  ("tmbim", lang2),
  ("a", ambiguous),
  ("Los", lang2),
  ("heats", ne),
]
```

Output: sentiment label

"negative"

# Example Vector

```
[  
("freshman", ne),  
("15", other),  
("empiezan", lang2),  
("hoy", lang2),  
("?", other),  
("?", other),  
]
```

->

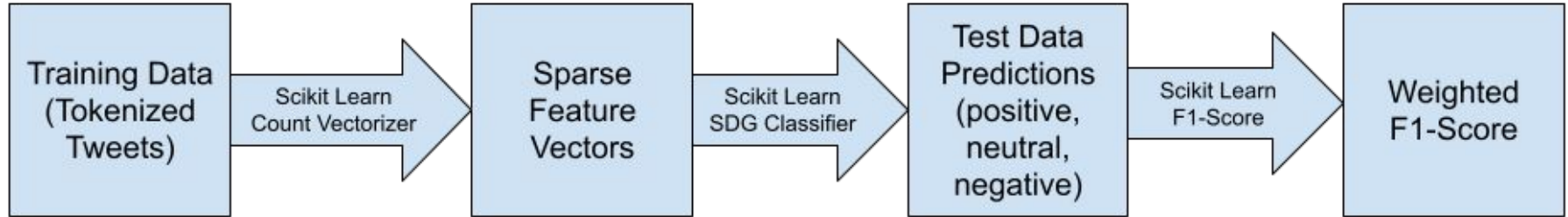
```
("<BOS>"): 1,  
("freshman"): 1,  
("15"): 1,  
("empiezan"): 1,  
("hoy"): 1,  
("?"): 2,  
("<EOS>"): 1
```

```
(ne): 1,  
(other): 3,  
(lang2): 2
```

```
("<BOS>", "freshman"): 1,  
("freshman", "15"): 1,  
("15", "empiezan"): 1,  
("empiezan", "hoy"): 1,  
("hoy", "?"): 1,  
("?", "?"): 1,  
("?", "<EOS>"): 1
```

```
(ne, other): 1,  
(other, lang2): 1,  
(lang2, lang2): 1,  
(lang2, other): 1,  
(other, other): 1
```

# Baseline System



# Results

System	Precision	Recall	F1
Random	0.1544	0.4551	0.4032
All Positive	0.1666	0.3333	0.3329
SGDClassifier	0.3865	0.4438	0.5075



# Discussion

- We experienced some variation in our F1 scores
  - 0.47 to 0.59
- Preprocessing the data (e.g. accounting for OOV tokens, lower-casing tweets, and standardizing encodings) did little to affect system accuracy.
- Model assigns most tweets a positive sentiment, very few negative -> more positive examples in our dataset.
  - Train Split: 6005 positive // 3974 neutral // 2023 negative
  - Dev Split: 1498 positive // 994 neutral // 506 negative
- F1 scores split:
  - Positive: 0.646449330300733
  - Neutral: 0.2448690728945506
  - Negative: 0.15654952076677314

# Discussion

- More context about tweet could be useful
- Multiple tweets are language lessons
- Dev set id #36, gold label is negative, model assigned neutral or positive



**Diana Estrada**

@HablaBienIngles



**At the drop of a dime: Instantáneamente sin vacilación .  
Don't test me, I'll leave you at the drop of a dime.**

9:35 PM · May 1, 2013 · Twitter Web Client

# Steps for the future

## Embeddings/Features

1. Multilingual (contextual) embeddings
2. Emoji embeddings e.g. emoji2vec from Eisner et al. (2016)

## Classifiers

3. Deep Averaging Network (DAN) from Mohit (2015)
4. Recurrent/seq based, (fine-tune BERT?)
5. Other classical ML

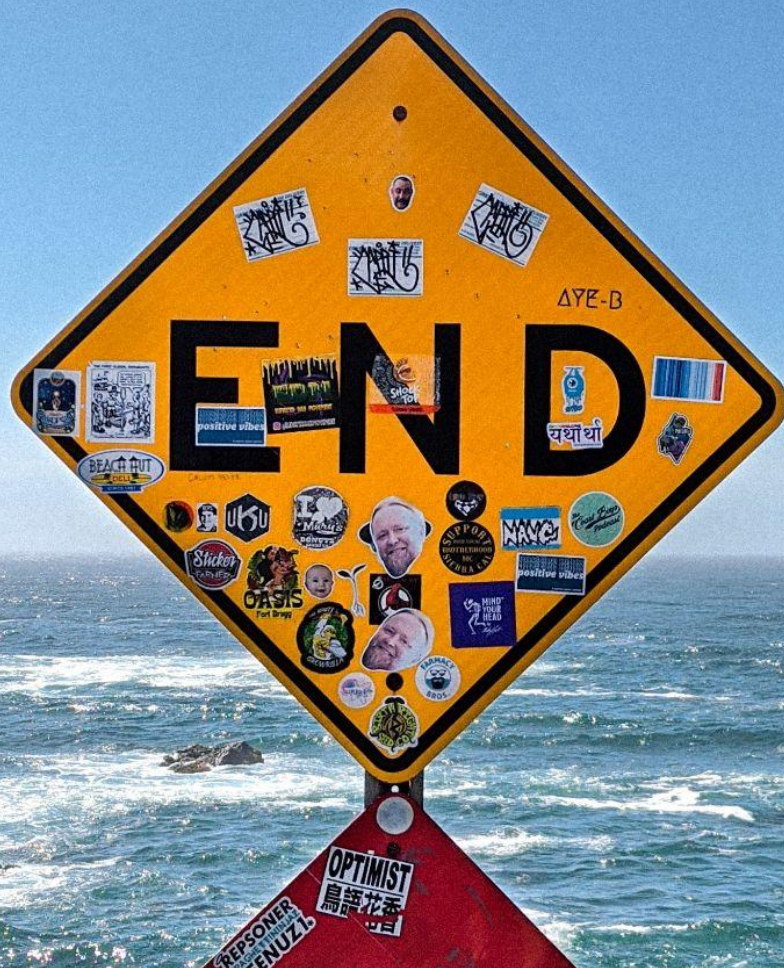
## Data

6. Deciding which “test” Spanglish dataset we should use.
  - a. SemEval set split: 12,000 train // 2,300 dev // 3,800 test (we have labels for 2,200 of them)
  - b. LinCE set split: 12,200 train // 1,860 dev // 4,700 test

# References

- Mohit Iyyer, et al. "Deep unordered composition rivals syntactic methods for text classification." *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)*. 2015.
- Gustavo Aguilar, Sudipta Kar, and Thamar Solorio. 2020. Lince: A centralized benchmark for linguistic code-switching evaluation.
- Francesco Barbieri, German Kruszewski, Francesco Ronzano, and Horacio Saggion. 2016. How cosmopolitan are emojis? exploring emojis usage and meaning over different languages with distributional semantics. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 531–535.
- Petra Kralj Novak, Jasmina Smailovic, Borut Sluban, and Igor Mozetic. 2015. Sentiment of emojis. *PloSone*, 10(12):e0144296.
- Parth Patwa, Gustavo Aguilar, Sudipta Kar, Suraj Pandey, Srinivas PYKL, Björn Gambäck, TanmoyChakraborty, Thamar Solorio, and Amitava Das. 2020. Semeval-2020 task 9: Overview of sentiment analysis of code-mixed tweets. In *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval-2020)*, Barcelona, Spain. Association for Computational Linguistics.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830
- Eisner, Ben, et al. "emoji2vec: Learning emoji representations from their description." *arXiv preprint arXiv:1609.08359* (2016).

Thank you!



PC: PP