Sentimiento Analysis

Classifying Affect in Mixed Language Tweets

Connor Boyle
Martin Horst
Nikitas Tampakis

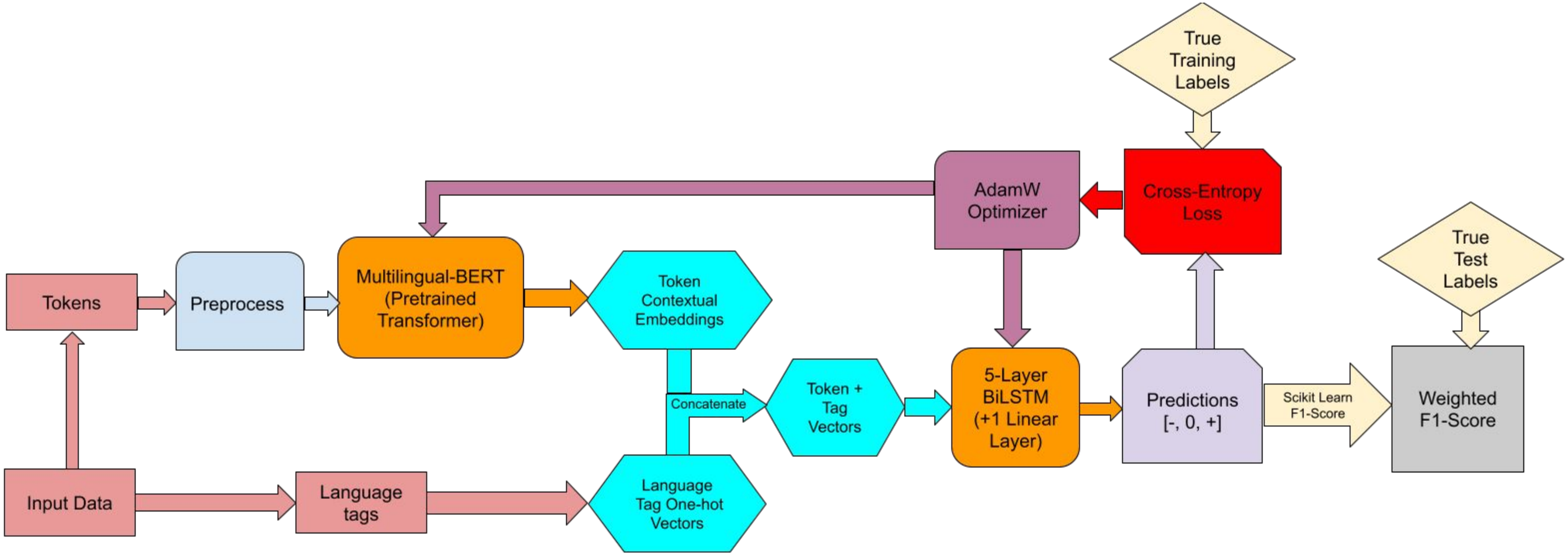
LING 573
Group 6
June 1, 2021

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

Parth Patwa[1*] Gustavo Aguilar[2*] Sudipta Kar[2] Suraj Pandey[3]
Srinivas PYKL[1] Björn Gambäck[4] Tanmoy Chakraborty[3]
Thamar Solorio[2] Amitava Das[5]

- Sentiment analysis on (Spanglish, Hinglish) tweets
- Categorical output: `[-, 0, +]`
- Adaptation task: Hinglish code mixed tweets

Final System Architecture



Example Input tweet - Hinglish

Original tweet +
lang tags

```
('@', 'O')
('PJkanojia', 'Hin')
('@', 'O')
('ssarwar2', 'Hin')
('Modi', 'Hin')
('ne', 'Hin')
('yagan', 'Hin')
('k', 'Eng')
('politi', 'Eng')
('…', 'O')
('https', 'Eng')
('//', 'O')
('t', 'Eng')
('.', 'O')
('co', 'Hin')
('/', 'O')
('ZVKWrYh8Rg', 'Eng')
('🚨😩', 'O')
```

->

Preprocessed tweet
(URLs, mentions, emojis)

```
('@USER', ['O', 'Hin'])
('@USER', ['O', 'Hin'])
('Modi', ['Hin'])
('ne', ['Hin'])
('yeh', ['Hin'])
('yagan', ['Hin'])
('k', ['Eng'])
('politi', ['Eng'])
('…', ['O'])
('HTTPURL', ['Eng', 'O', 'Eng',
('revolving light', ['O'])
('weary', ['O'])
```

All tags: ['lang1', 'lang2', 'ne', 'ambiguous', 'unk', 'other', 'mixed', 'fw', 'Eng', 'O', 'Hin']

->

Dense contextual embeddings

One-hot language vectors

```
[0.34, 0.12, 0.87, …, 0, 1, 1]
[-0.27, 1.32, 0.23, …, 0, 1, 1]
[1.83, -3.12, 2.15, …, 0, 1, 0]
[3.41, -2.75, 0.19, …, 0, 1, 0]
[0.23, 1.83, -2.37, …, 0, 1, 0]
[-2.34, 4.21, 8.52, …, 1, 0, 0]
[2.83, 1.98, 5.22, …, 1, 0, 0]
[-1.32, 4.83, 9.63, …, 0, 0, 1]
[2.80, 2.64, 2.71, …, 0, 2, 1]
[1.29, 4.74, 2.76, …, 0, 0, 1]
[-2.01, 6.32, 5.01, …, 0, 0, 1]
```

Results - D2/D3

System	**Positive** Precision/Recall F1	**Neutral** Precision/Recall F1	**Negative** Precision/Recall F1	**Avg. F1**
Spanglish Linear SVM (D2)	0.5386 0.7499 0.6252	0.3689 0.2628 0.2971	0.4157 0.1577 0.2233	0.4486
Hinglish Linear SVM (D2)	0.6500 0.6448 0.6432	0.5372 0.4505 0.4838	0.5841 0.6930 0.6302	0.5794
M-BERT - Spanglish (D3)	0.5981 0.6348 0.6159	0.4028 0.4296 0.4158	0.4253 0.2925 0.3466	0.5041
M-BERT - Hinglish (D3)	0.6568 0.6782 0.6673	0.5401 0.4832 0.5101	0.6059 0.6652 0.6342	0.5984

Results - D4 Dev/Eval

System	**Positive** Precision/Recall F1	**Neutral** Precision/Recall F1	**Negative** Precision/Recall F1	**Avg. F1**
M-BERT+BiLSTMClassifier (Spanglish) Dev	0.5767 0.7576 0.6549	0.3791 0.2052 0.2663	0.3963 0.3854 0.3908	0.4815
M-BERT+BiLSTMClassifier (Hinglish) Dev	0.6790 0.6721 0.6755	0.5311 0.5000 0.5151	0.6025 0.6539 0.6272	0.6008

Results - D4 Hinglish Comparison

System	**Positive** Precision/Recall F1	**Neutral** Precision/Recall F1	**Negative** Precision/Recall F1	**Avg. F1**
Baseline	0.728 0.688 0.707	0.562 0.602 0.581	0.691 0.674 0.683	0.654
M-BERT+BiLSTMClassifier (Hinglish) Test	**0.7774 0.7370 0.7567**	**0.6013 0.5936 0.5974**	**0.6894 0.7400 0.7138**	**0.6854**
Top Performer	0.843 0.760 0.799	0.652 0.731 0.689	0.785 0.754 0.769	.750

Results - D4 Spanglish* Comparison

System	Avg. F1
Baseline (M-BERT)	0.5643
M-BERT+BiLSTMClassifier	**0.4849**
Top Performer	0.6126

*separate task; different dataset from original SentiMix

Discussion

a. Issues
    i. Freezing M-BERT layers degraded model performance more than it helped prevent catastrophic forgetting
    ii. Adding dropout to the last LSTM layer degraded model performance more than it helped with overfitting
    iii. Incorporating language tags didn't seem to have much impact
    iv. Skewed data sets = more positive classifications, 55% in Spanglish data positive, 30% neutral, 15% negative
    v. Preprocessing data seemed irrelevant for Spanglish, but helped with Hinglish
b. Successes
    i. End to end NN system, check
    ii. Large file/large model handling
    iii. Beat Hinglish baseline

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.

Chiorrini, Andrea & Diamantini, Claudia & Mircoli, Alex & Potena, Domenico. (2021). Emotion and sentiment analysis of tweets using BERT.

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, Bj ̈orn Gamb ̈ack, 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).

Harel-Canada, Fabrice. emoji_translate Python Package, https://github.com/fabriceyhc/emoji_translate

Thank you!

[PC: PP]