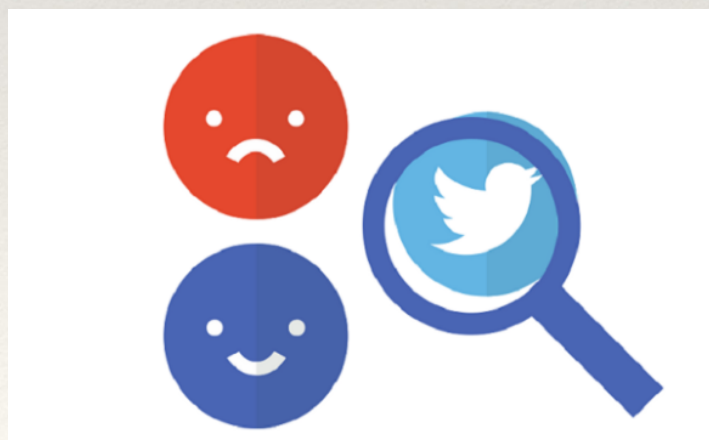# Generate Sentiment Polarity of Tweets Using LSTM with Word2Vec/GloVe

G-Langlanucher

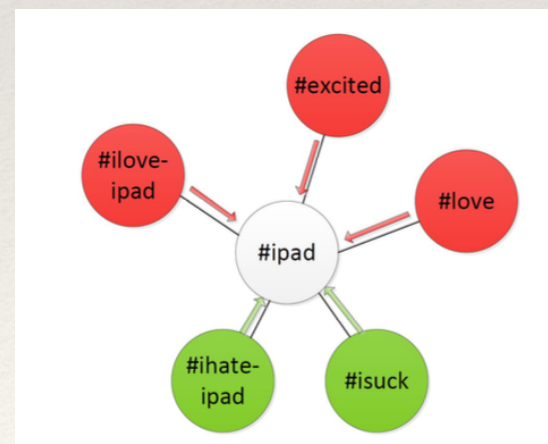*Wei Yao*

CMPT 413

Term Project

Problem

❖ Previous: Classification of Tweets Based on Automated Twitter Hashtag Generated from Sentiment Analysis Model

❖ Now: The phase I - Sentiment Analysis of Tweets: classify them into two groups {Pos, Neg}


❖ Phase I

❖ Phase II

Data

	target	ids	date	flag	user	text
0	NEGATIVE	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	NEGATIVE	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
2	NEGATIVE	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
3	NEGATIVE	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
4	NEGATIVE	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all....
5	NEGATIVE	1467811372	Mon Apr 06 22:20:00 PDT 2009	NO_QUERY	joy_wolf	@Kwesidei not the whole crew
6	NEGATIVE	1467811592	Mon Apr 06 22:20:03 PDT 2009	NO_QUERY	mybirch	Need a hug
7	NEGATIVE	1467811594	Mon Apr 06 22:20:03 PDT 2009	NO_QUERY	coZZ	@LOLTrish hey long time no see! Yes.. Rains a...
8	NEGATIVE	1467811795	Mon Apr 06 22:20:05 PDT 2009	NO_QUERY	2Hood4Hollywood	@Tatiana_K nope they didn't have it
9	NEGATIVE	1467812025	Mon Apr 06 22:20:09 PDT 2009	NO_QUERY	mimismo	@twittera que me muera ?

Data

- ❖ Data Distribution

Labels: ['NEGATIVE', 'POSITIVE']
Amounts: [800000, 800000]
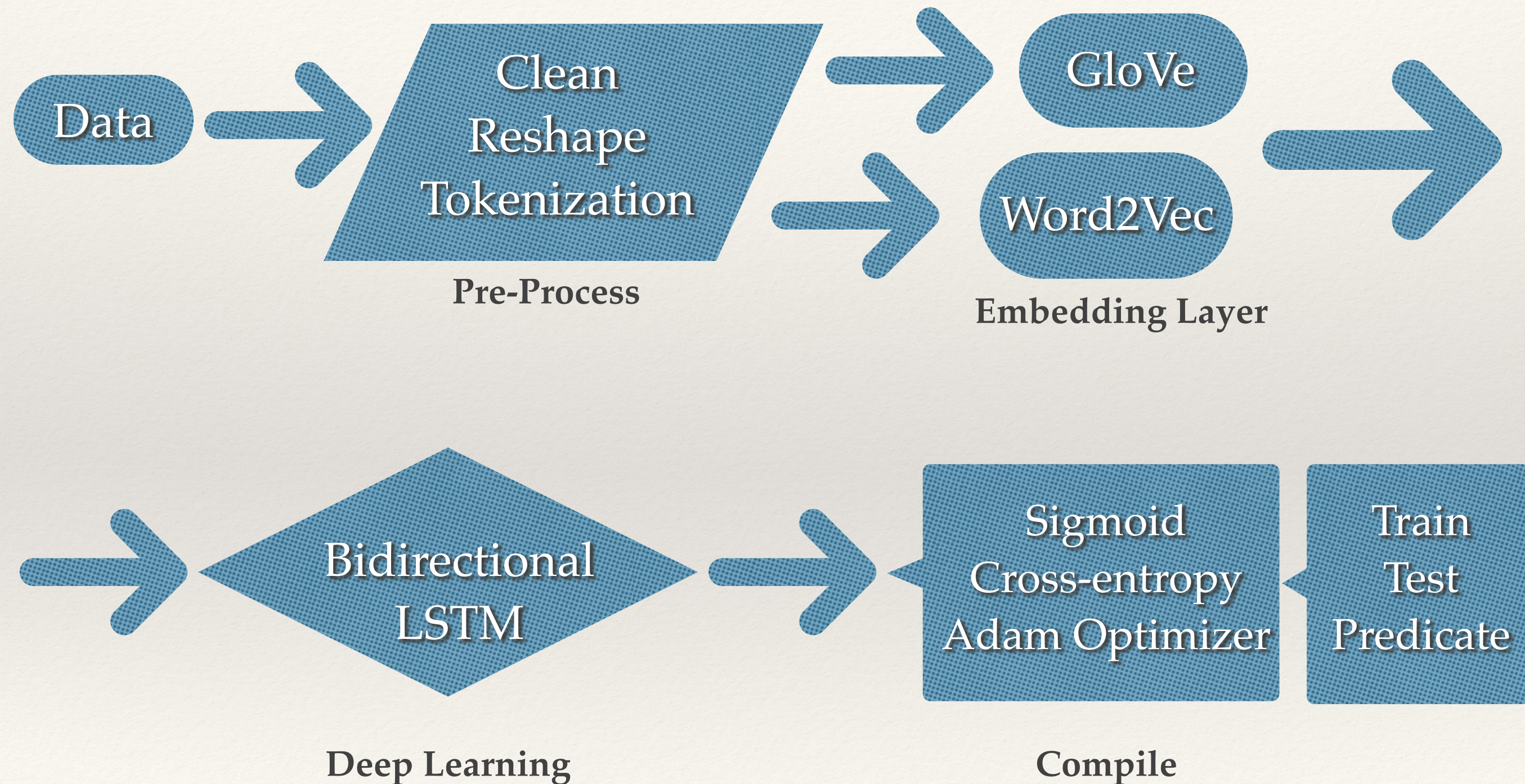
Sentiment Labels

NEGATIVE 50.0%
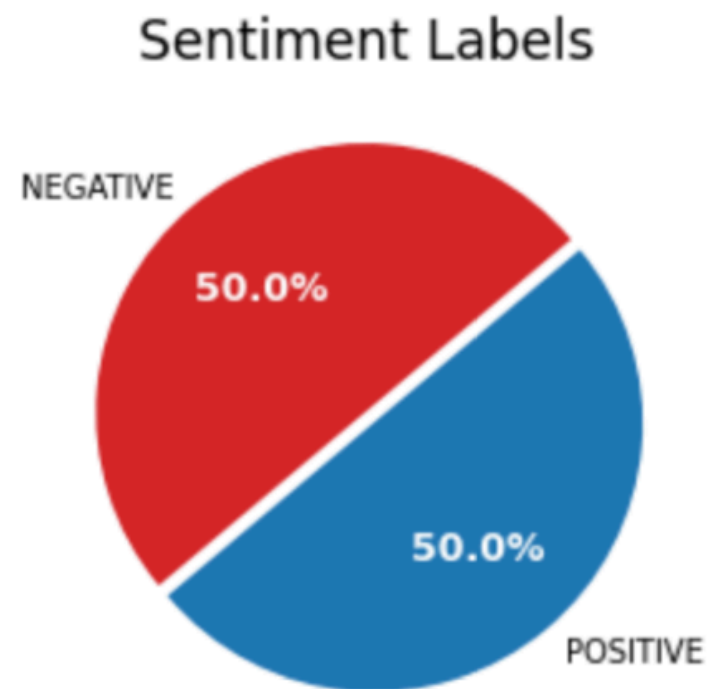
POSITIVE 50.0%
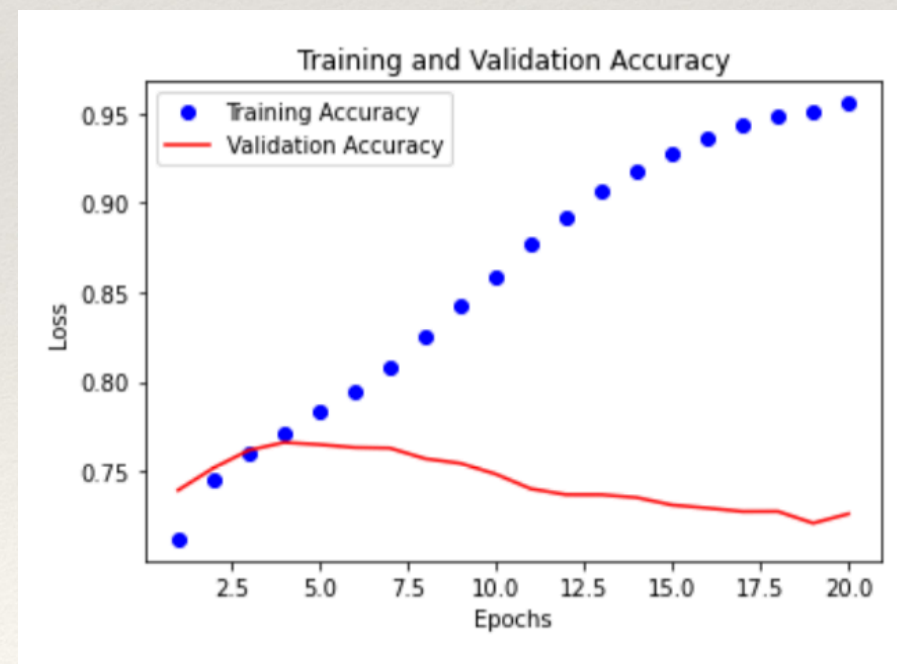
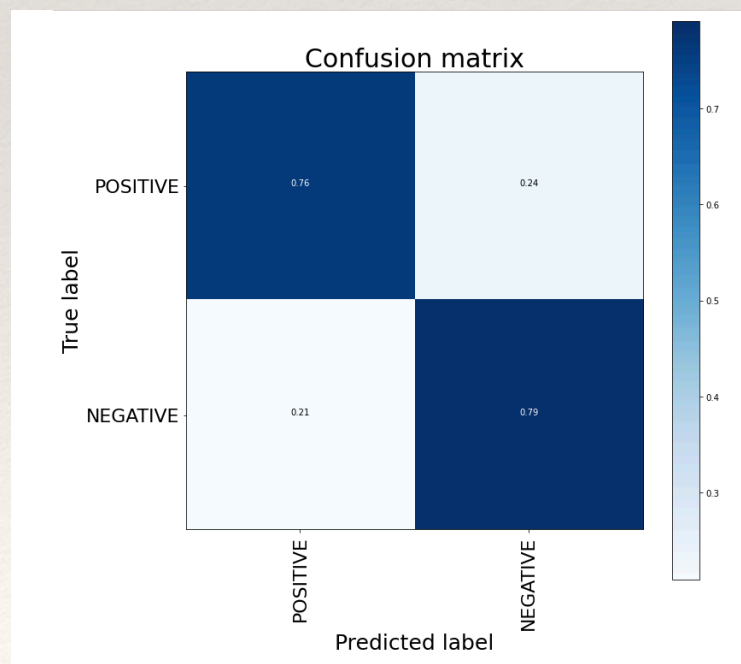- ❖ Accurate
- ❖ Balanced
- ❖ Cleaned

Result

❖ Capability

```
predict("NLP is so interesting that I am getting obsessed")

{'label': 'POSITIVE',
 'score': 0.9143790602684021,
 'elapsed_time': 0.23479557037353516}
```

❖ Performance:

Confusion matrix

Training and Validation Accuracy

❖ Word2Vec   vs   ❖ GloVe

Analysis

❖ Why Word2Vec isn't Good?

Conclusion

❖ Accuracy: over 90%

❖ GloVe is much better than Word2Vec

❖ Future: Overfitting - Dropout ?

Accomplish our original goal?

Reference

B. Jansen, M. Zhang, K. Sobel, A. Chowdury. The Commerical Impact of Social Mediating Tech- nologies: Micro-blogging as Online Word-of- Mouth Branding, 2009.

B. Pang and L. Lee. "Opinion Mining and Sen- timent Analysis" in Foundations and Trends in In- formation Retrieval, 2008.

B. Pang, L. Lee, S. Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques, 2002.

P. D. Turney, Thumbs up or thumbs down ? : semantic orientation applied to unsupervised clas- sification of reviews, in ACL '02 : Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417–424, Morris- town, NJ, USA, 2001, Association for Computa- tional Linguistics.

Thanks !!!