Contribution to research new models of knowledge extraction on BigData systems

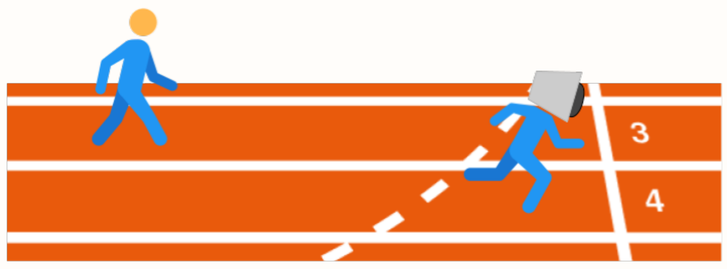
Héctor Cerezo-Costas, Advisor: F.Javier González-Castaño[1]

[1]AtlantTIC. Department of Telematics Engineering, University of Vigo

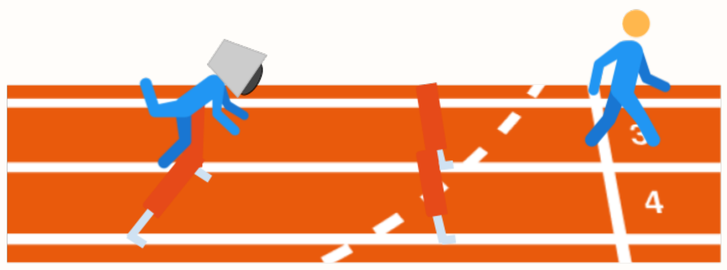
Motivation

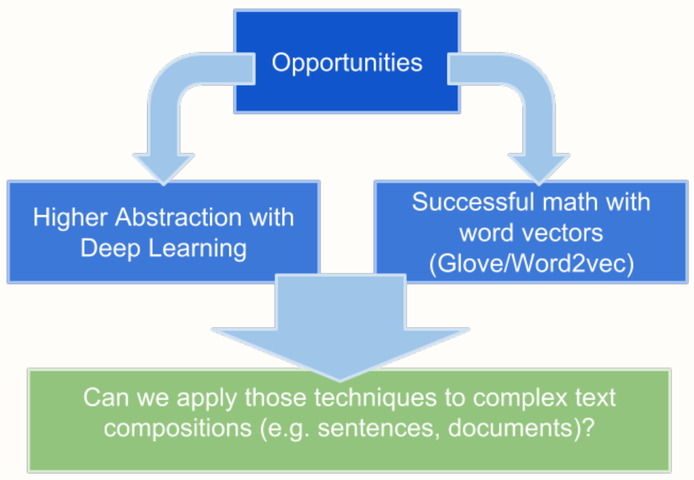
Need to improve Natural Language Understanding (NLU) by computers

Computers are faster at processing web-scale data

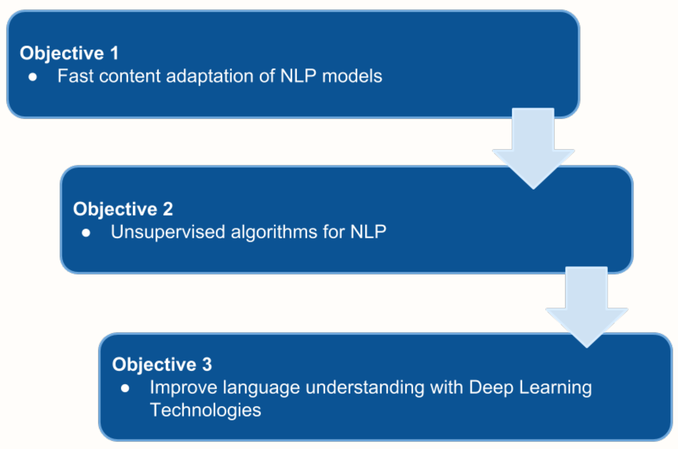


In complex NLP tasks humans have unparalleled performance.





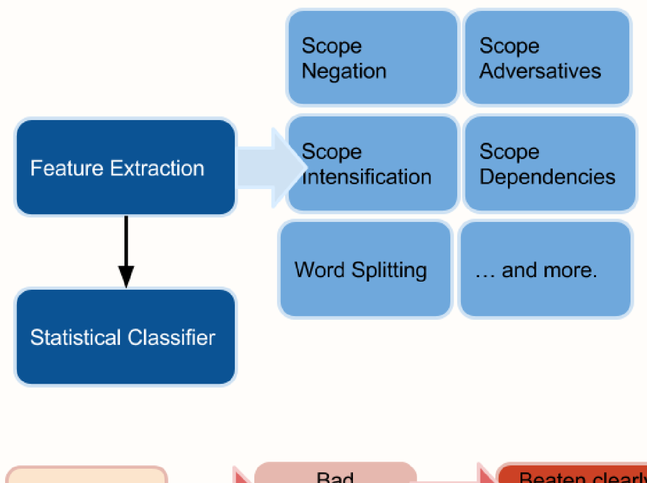
Thesis Objectives



Results

Sentiment Analysis

Traditional approach:



Achievements:

SemEval 2015 Competition [1]:

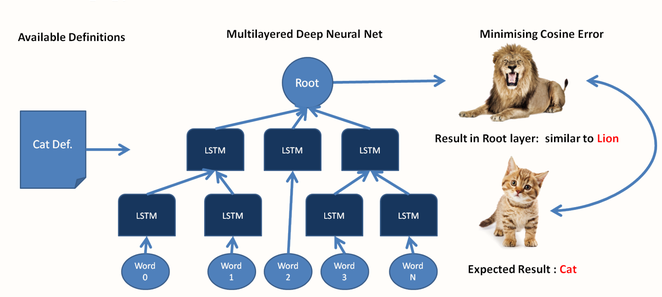
- 16th position out of 40 competitors in both sarcasm and regular 2015 datasets.
- 1st position in 2014 Tweet Sarcasm dataset.

TASS 2015 Spanish competition [2]

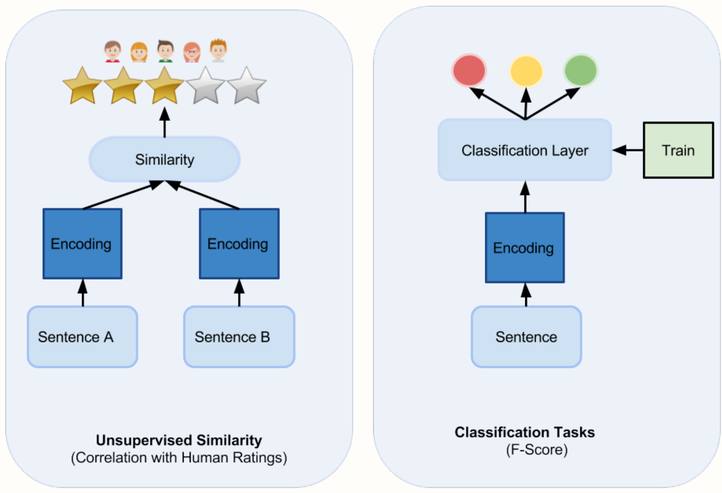
- 2nd position sentiment analysis in Twitter.

Learning Semantic Sentence Representations

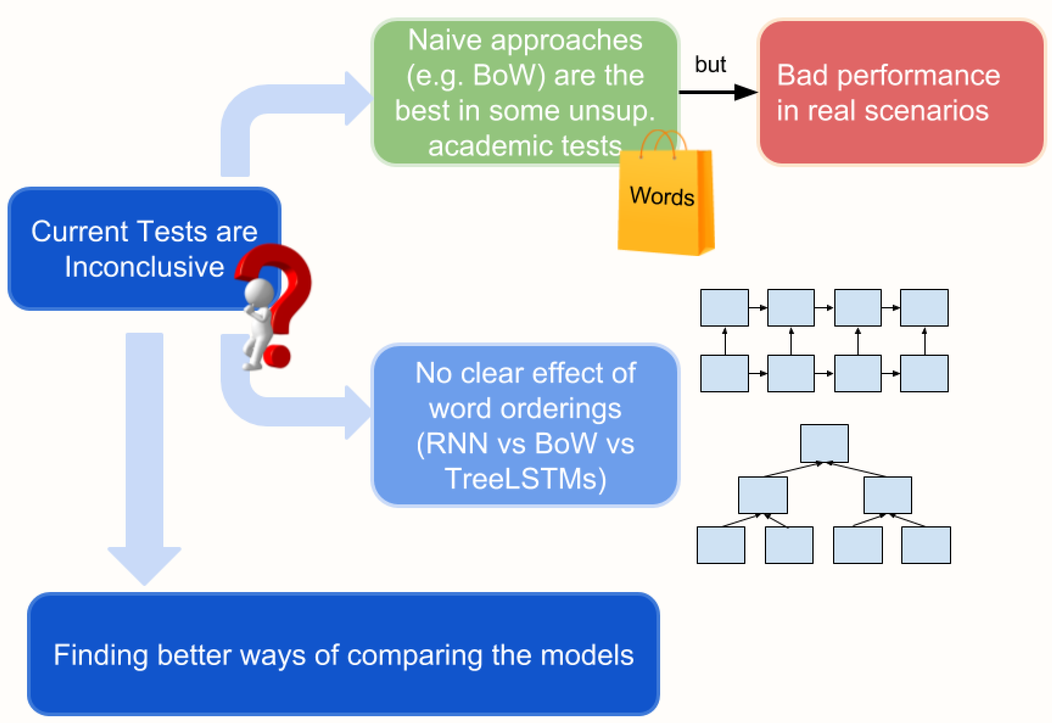
Goal: learning sentence abstraction with dictionary definitions and dynamic graph networks (e.g. dependency graphs [3]).



Performance comparison with naive Bag of Words (BoW) and SOTA models in supervised and unsupervised benchmarks [4]:



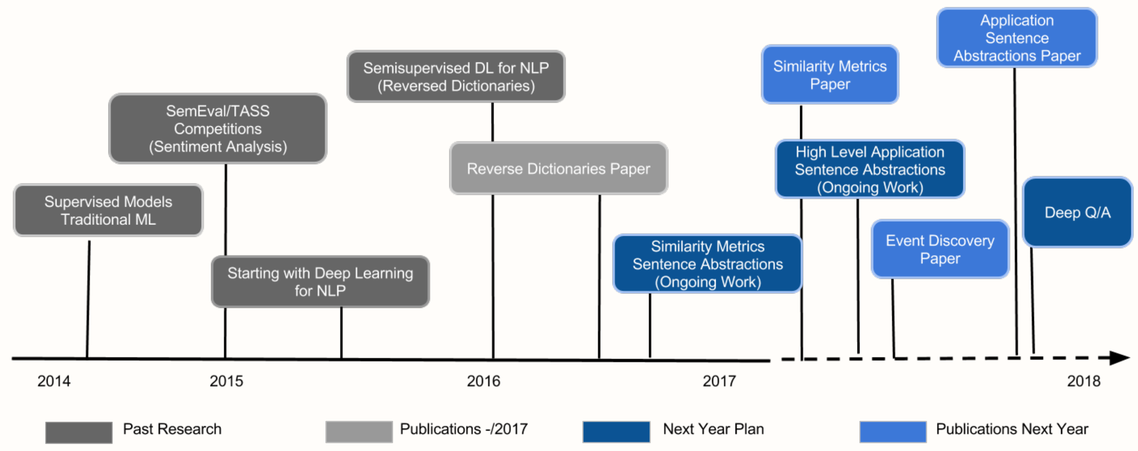
Critical Thinking



Still far from human performance but there is much room for improvement:

- Dynamic Network Graphs seem just the right idea for encoding sentences (recent paper from Google Deepmind [5])
- Better ordering of the semantic codification is possible (Continuous Variational Autoencoders [6]).

Research Plan (Next Year)



References

[1] H. Cerezo-Costas, and D. Celix-Salgado, (2015, June). Gradiant-analytics: Training Polarity Shifters with CRFs for Message Level Polarity Detection. *In Proc. of the 9th Int. Workshop on Semantic Evaluation (SemEval 2015) (pp. 539-544).*

[2] T. Alvarez-López, J. Juncal-Martınez, M. Fernández-Gavilanes, E. Costa-Montenegro, F.G. González-Castano, H. Cerezo-Costas, and D. Celix-Salgado GTI-Gradiant: A Hybrid Approach for Sentiment Analysis in Twitter. *In Proc. of TASS 2015*

[3] K. S. Tai, R. Socher and C.D. Manning, (2015). Improved Semantic Representations from Tree-Structured Long Short-Term Memory Networks.

[4] H. Cerezo-Costas, M. Martín Vicente and F-J. González Castaño. Tree LSTMs for Learning Sentence Representations (Under Review)

[5] D. Yogatama, P. Blunsom, C. Dyer, E. Grefenstette and W. Ling (2016). Learning to compose words into sentences with reinforcement learning.

[6] S.R. Bowman, L. Vilnis, O. Vinyals, A.M Dai, R. Jozefowicz, and S. Bengio, (2015). Generating Sentences from a Continuous Space.

Workshop on Monitoring PhD Student Progress, 22-23 June 2017, Vigo, Spain