NLP News and Trendy Topics

Adrian M.P. Brasoveanu, MT
Guilherme Rodrigues, SWC

Introduction

Rada Mihalcea
@radamihalcea

Follow

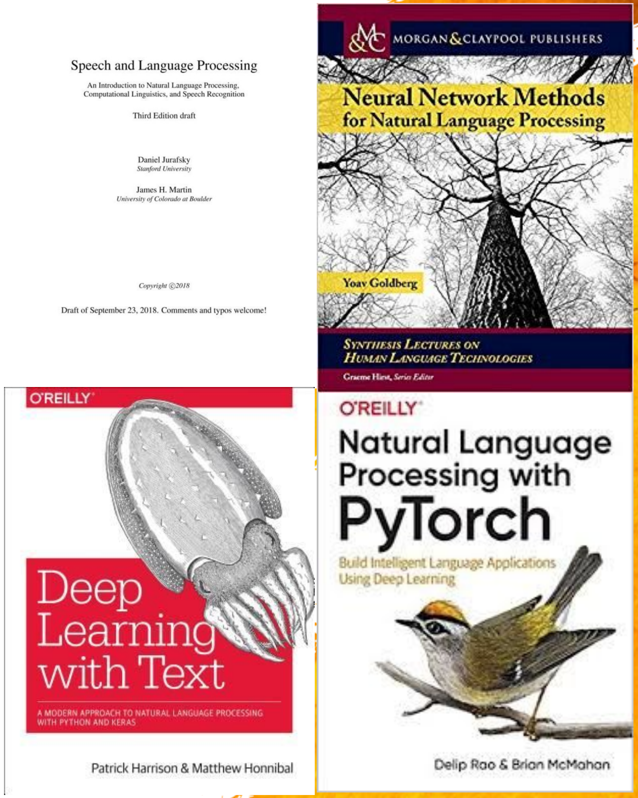
How words change meaning over time:

“Baseline” (in NLP)
1998: random choice
2005: majority class
2012: bag of words
2019: LSTM with attention & pre-trained embeddings

2:45 PM - 11 May 2019

212 Retweets **1,226** Likes

10 212 1.2K



New Frameworks

- **PyTorch is where it's at!**
- **Pythia**
  - a PyTorch framework for **bridging the gap**
  - DL for **Text** and **Computer Vision**
  - great for **VQA** challenges
  - https://github.com/facebookresearch/pythia
- **transfer-NLP**
  - **Transfer code** and **Transfer Learning**
  - Based on PyTorch and Delip Rao book
  - https://github.com/feedly/transfer-nlp

Large-scale data sets

- **DL** needs large data sets
  - 10k to 1M examples or more
- **DecaNLP –** MTL as QA
- **GLUE** – MTL benchmark for NLU
- **SquAD 2.0** – Stanford QA
- **SWAG** – Situations with Adversarial Generation
- **VGR** – Visual Commonsense Reasoning
- **Gab.ai** – Social Media posts
- **NarrativeQA** – Reading Comprehension

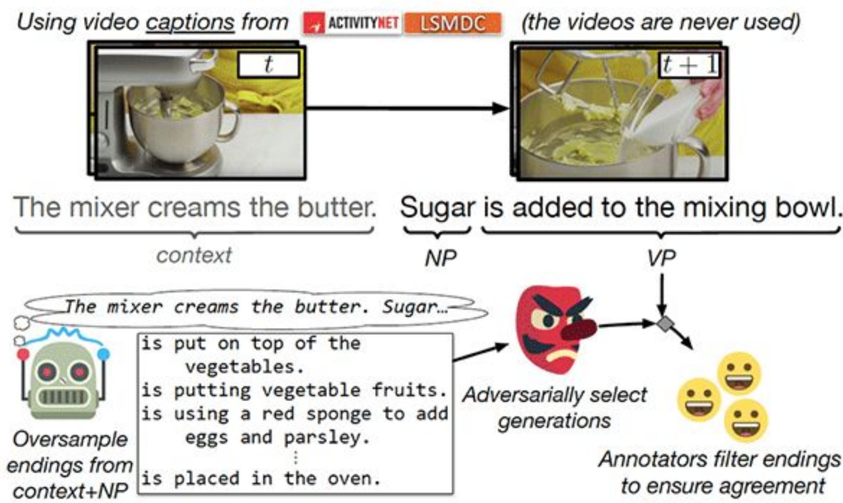
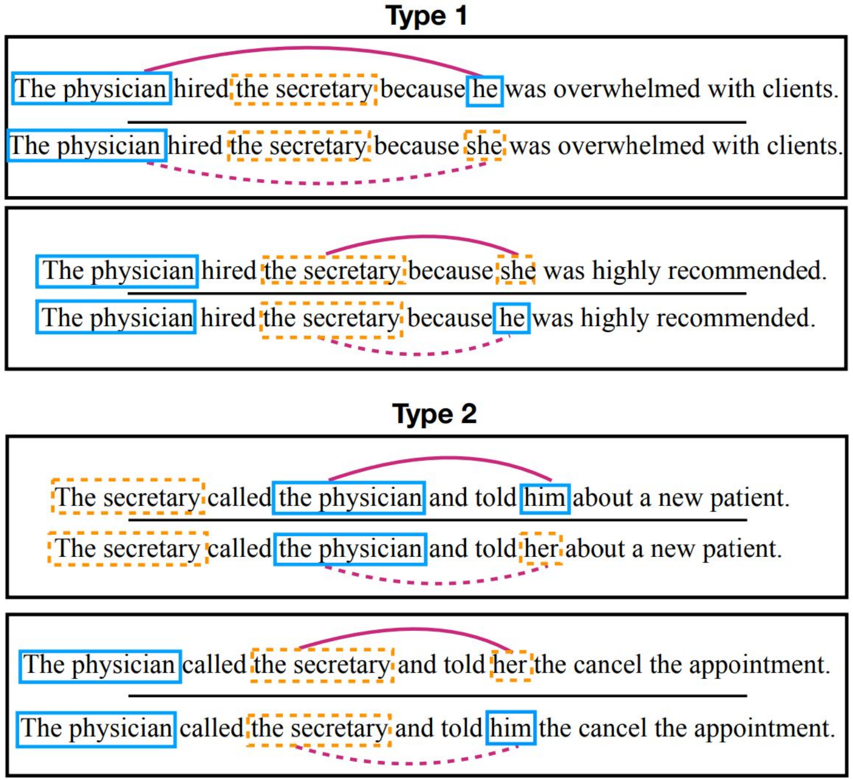


Figure 1: Overview of the data collection process.

Bias

- Gender and representation bias
- Unconscious bias
- Aim for balanced gender
- IBM's AI Fairness 360
  - 10 fairness SoA algorithms
- Relational Inductive biases in Graphs
  - Bias is in the networks
  - https://arxiv.org/pdf/1806.01261.pdf
- Gender and Resource Co-reference Bias
  - https://arxiv.org/pdf/1804.06876.pdf

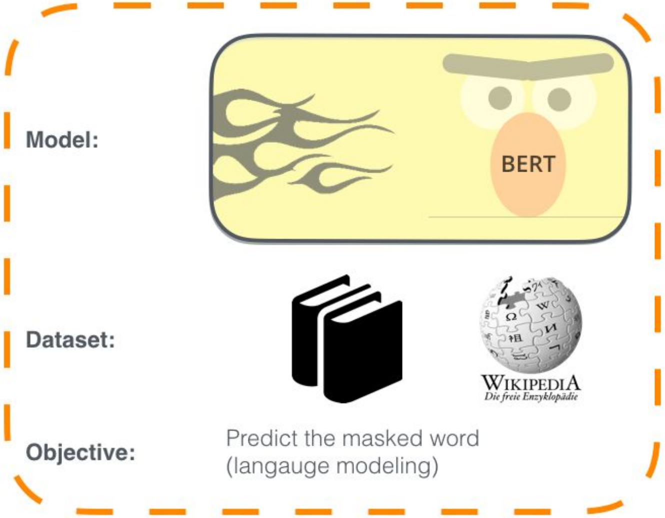


BERT and Friends

1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

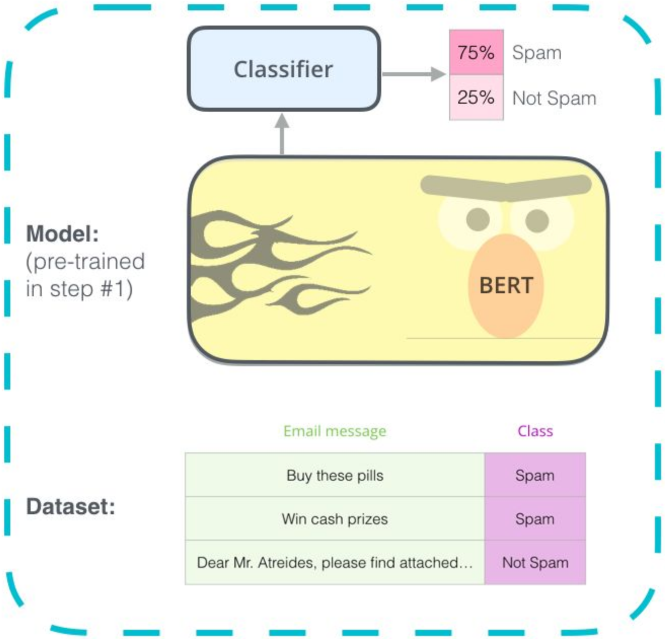
The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

Semi-supervised Learning Step



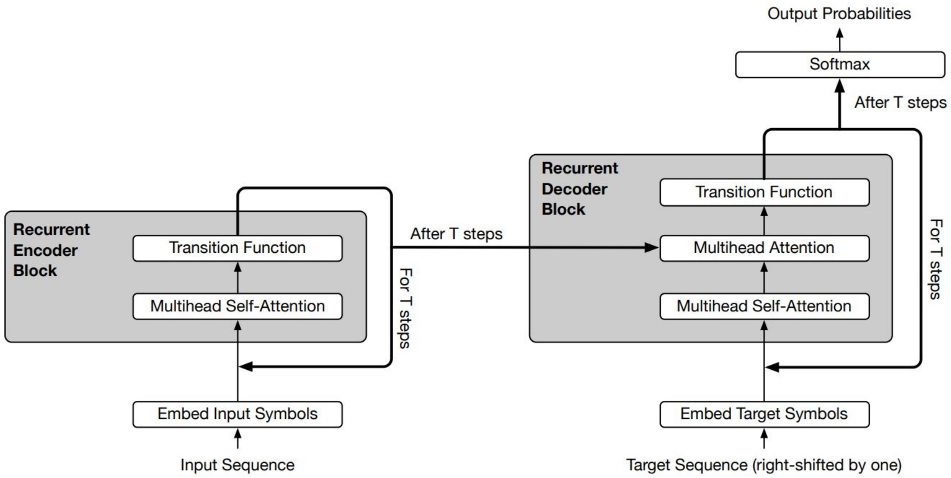
2 - Supervised training on a specific task with a labeled dataset.

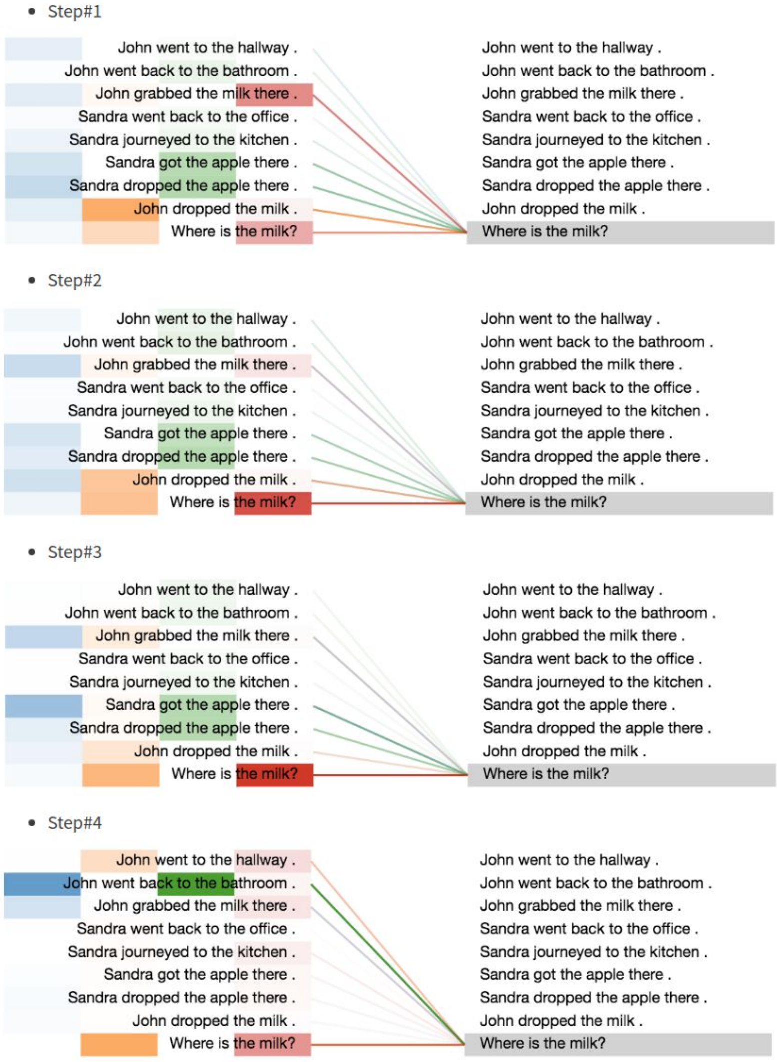
Supervised Learning Step



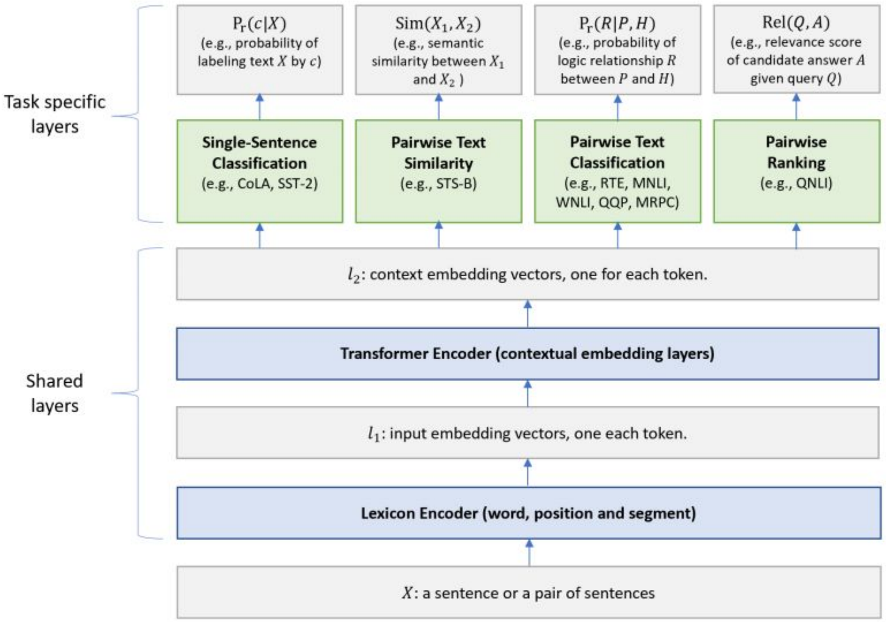
Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached…	Not Spam

Universal Transformers



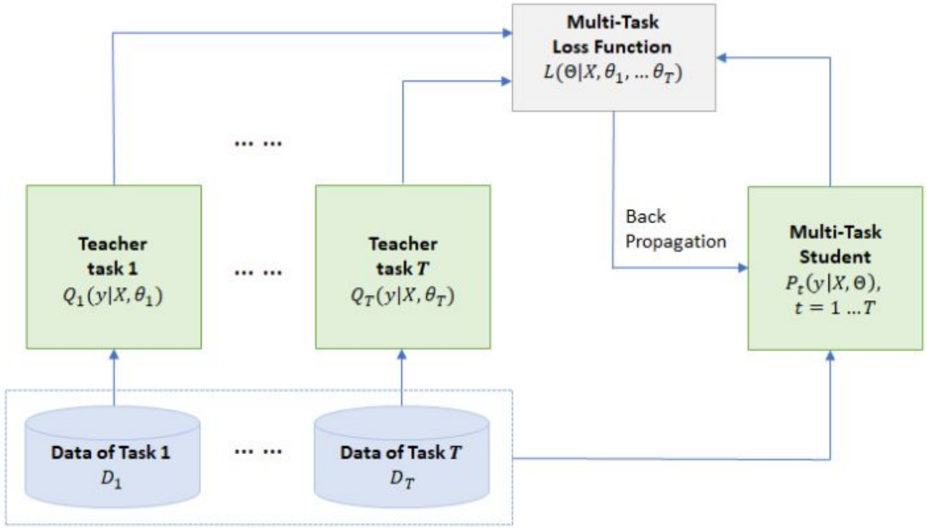


Microsoft MT-DNN



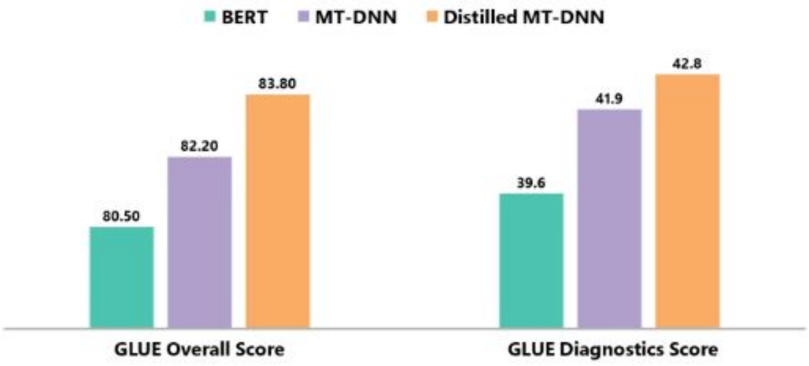
Multi-Task Deep Neural Networks for Natural Language Understanding, https://arxiv.org/abs/1901.11504

Knowledge Distillation



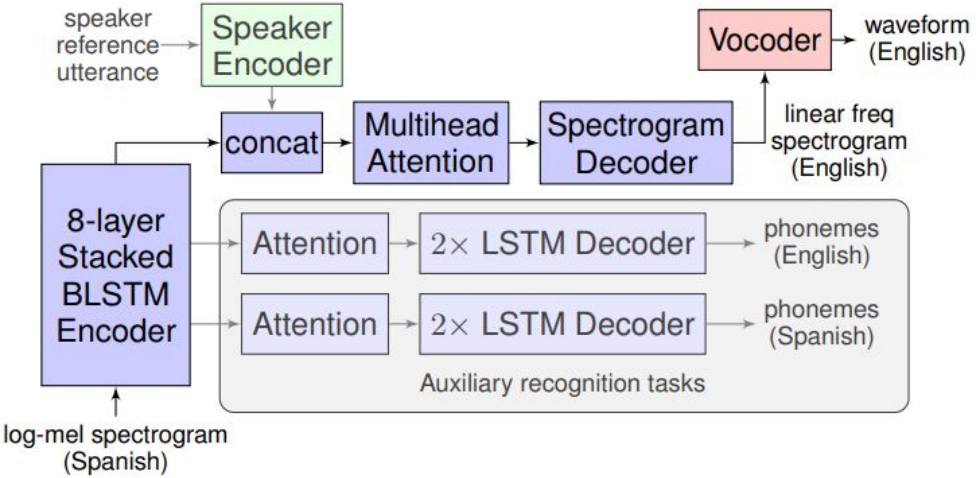
Improving Multi-Task Deep Neural Networks via Knowledge Distillation for Natural Language Understanding, https://www.microsoft.com/en-us/research/publication/improving-multi-task-deep-neural-networks-via-knowledge-distillation-for-natural-language-understanding/

Knowledge Distillation



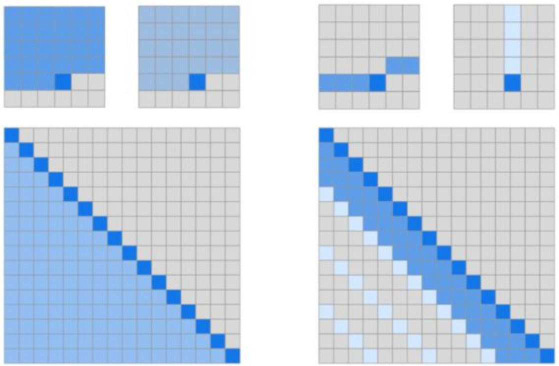
Microsoft makes Google’s BERT NLP model better,
https://venturebeat.com/2019/05/16/microsoft-makes-googles-bert-nlp-model-better/

Google AI 'Translatotron'



Direct speech-to-speech translation with a sequence-to-sequence model https://arxiv.org/abs/1904.06037

OpenAI Sparse Transformers



(a) Transformer (b) Sparse Transformer (strided) (c) Sparse Transformer (fixed)

Generative Modeling with Sparse Transformers https://openai.com/blog/sparse-transformer/
Generating Long Sequences with Sparse Transformers https://arxiv.org/abs/1904.10509

OpenAI

Multi-task learning (MTL) has led to successes in many applications of machine learning, from natural language processing and speech recognition to computer vision and drug discovery. This article aims to give a general overview of MTL, particularly in deep neural networks. It introduces the two most common methods for MTL in Deep Learning, gives an overview of the literature, and discusses recent advances. The article then briefly considers the advantages and challenges of using MTL in combination with regularization and classification. Finally, a brief overview of the various approaches to MTL is provided, before examining more technical questions that arise from the use of MTL.

Talk with Transformer https://talktotransformer.com/

Thanks!

Any questions?

You can find us online :)