

# NLP Resource Creation and Enrichment using Deep Learning

Kevin Patel

Guided by:

Prof. Shivaram Kalyanakrishnan

Prof. Pushpak Bhattacharyya

April 1, 2016

# Natural Language Processing (NLP)

- Different NLP tasks

# Natural Language Processing (NLP)

- Different NLP tasks
  - Machine Translation

# Natural Language Processing (NLP)

- Different NLP tasks
  - Machine Translation
  - Information Retrieval

# Natural Language Processing (NLP)

- Different NLP tasks
  - Machine Translation
  - Information Retrieval
  - Dialogue Systems

# Natural Language Processing (NLP)

- Different NLP tasks
  - Machine Translation
  - Information Retrieval
  - Dialogue Systems
- Resource driven

# Natural Language Processing (NLP)

- Different NLP tasks
  - Machine Translation
  - Information Retrieval
  - Dialogue Systems
- Resource driven
  - WordNet
  - IndoWordNet, EuroWordNet
  - FrameNet
  - Parallel Corpora
  - Sense-annotated Corpora
  - Pretrained vector repositories

- Deep Learning



- Deep Learning
  - Traffic Sign Recognition
  - Handwriting recognition (English and other languages such as Chinese, Arabic)
  - Breast Cancer Prognosis

- Deep Learning
  - Traffic Sign Recognition
  - Handwriting recognition (English and other languages such as Chinese, Arabic)
  - Breast Cancer Prognosis
- Deep Learning in NLP

- Deep Learning
  - Traffic Sign Recognition
  - Handwriting recognition (English and other languages such as Chinese, Arabic)
  - Breast Cancer Prognosis
- Deep Learning in NLP
  - *SENNA* : POS, chunking, NER, SRL

- Deep Learning
  - Traffic Sign Recognition
  - Handwriting recognition (English and other languages such as Chinese, Arabic)
  - Breast Cancer Prognosis
- Deep Learning in NLP
  - *SENNA* : POS, chunking, NER, SRL
  - *DSSM* : web-search ranking, contextual searching, advertising relevance

- Deep Learning
  - Traffic Sign Recognition
  - Handwriting recognition (English and other languages such as Chinese, Arabic)
  - Breast Cancer Prognosis
- Deep Learning in NLP
  - *SENNA* : POS, chunking, NER, SRL
  - *DSSM* : web-search ranking, contextual searching, advertising relevance
- A crucial development - word vectors (word embeddings, word representations, etc.)

## Enriching existing resources

### Automatic linking of Synsets

- Within a language specific wordnet
  - Cross lingual
- 

### Refining pretrained vector repositories

- Detection and removal of non-specific vectors
- Estimating appropriate task specific representation for out-of-vocabulary words

## Creating new resources

### Creating vector representations of complex lexical entities such as

- Synsets
  - Phrases
  - Sentences
  - Question/Answer pairs
- 

### Investigating compositional and non-compositional methods of creating vectors

# Thank You

Questions?