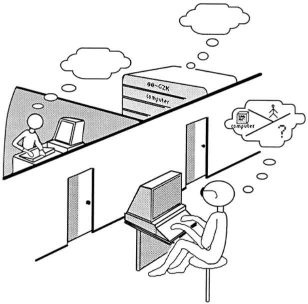
Deep Natural Language Processing

Yogesh Kulkarni

Introduction

Turing Test

A human judge engages in a natural language conversation with two other parties, one a human and the other a machine; if the judge cannot reliably tell which is which, then the machine is said to pass the test.



Use of Deep Learning in Natural Language Processing is helping beat the Turing test, a mark of intelligence, ie Artifical Intelligence.

NLP-AI

- Natural Language Processing (Understanding-Generation) is an AI-Complete problem
- Most difficult AI problems are informally known as AI-complete or AI-hard
- To call a problem AI-complete reflects an attitude that it would not be solved by a simple specific algorithm.

(Ref: https://en.wikipedia.org/wiki/AI-complete)

NLP Tasks

- Part Of Speech Tagging: Assign part-of-speech to each word.
- Named Entity Recognition: Recognize people, places, etc. in a sentence.
- Language Modeling: Predict next word, Generate natural sentences.
- Translation: Translate a sentence into another language.
- Summarization: Summarize a paragraph in new words.

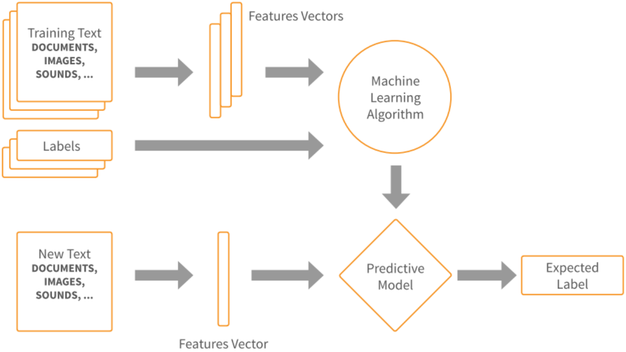
(Ref: How Deep Learning Quietly Revolutionized NLP - Lukasz Kaiser, Google Brain)

Let's discuss

- What is Deep Learning?
- Why is it important?
- How to apply Deep Learning to Natural Language Processing?

(Ref (next few slides): Deep Learning for Natural Langauge Processing - Sihem Romdhani)

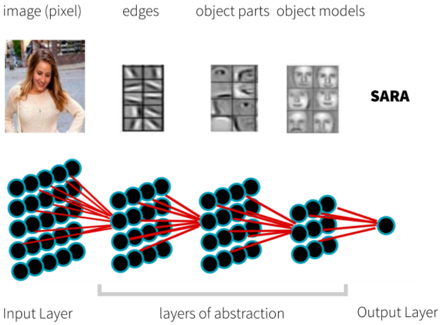
Machine Learning



(Ref: Deep Learning for Natural Language Processing - Sihem Romdhani)

Hand crafted features are needed in Machine Learning. E.g. for Spam Detection, features could be presennce of BIG $ amounts, FROM country, etc.

Deep Learning

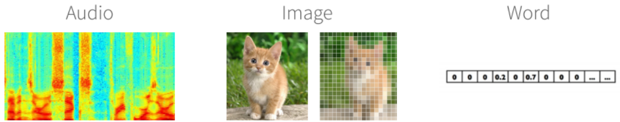


(Ref: Deep Learning for Natural Language Processing - Sihem Romdhani)

Hand crafted features are NOT needed in Deep Learning. E.g. for Object Detection, CNNs are come up with own features like, edges, parts, etc.

Need numbers not words!!

How can we represent words?



(Ref: Deep Learning for Natural Language Processing - Sihem Romdhani)

Words, sentences, paragraphs, pages, documents, all need to represented by numbers, to be able to fed to ML/DL algorithm.

Traditional methods to covert text to numeric

With Vocabulary as the full set representing the whole corpus.

- Bag-of-Word: Count of occurrences of a word in a document
- TF-IDF: Measures importance of a word in a document relative to the corpus

Challenges in traditional methods of encoding

With Vocabulary as the full set representing the whole corpus.

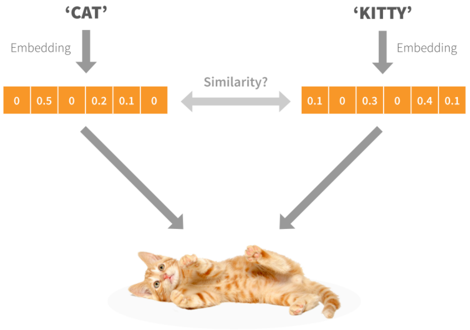
- Sparse inputs
- Context lost in encoding

```
A quiet crowd entered the historic church
                    !=
A historic crowd entered the quiet church
```

(Ref: Deep Learning for Natural Language Processing - Bargava, Amit Kapoor)

Embeddings

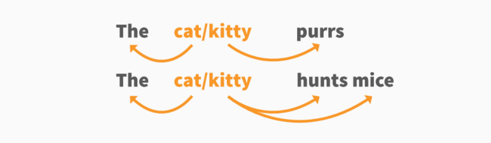
It just can not be any set of numbers representing words. Similar words should have similar numbers.



(Ref: Deep Learning for Natural Language Processing - Sihem Romdhani)

How to get good Embeddings?

Context defines similarity

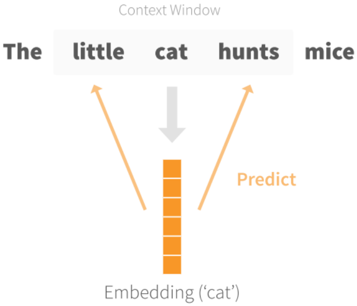


(Ref: Deep Learning for Natural Language Processing - Sihem Romdhani)

Infer the meaning of words from the company they keep

Word to Vector

Context defines similarity

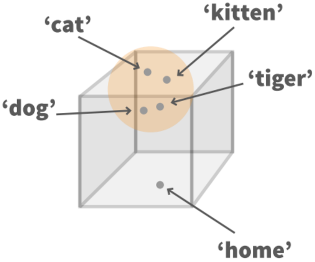


(Ref: Deep Learning for Natural Language Processing - Sihem Romdhani)

Infer the meaning of words from the company they keep

Word to Vector Examples

Similar words are nearby

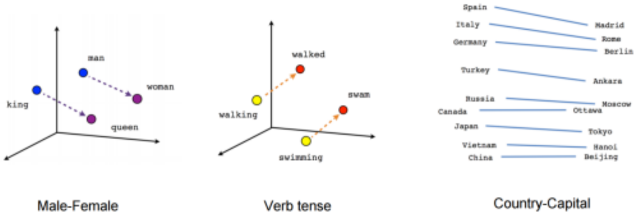


(Ref: Deep Learning for Natural Language Processing - Sihem Romdhani)

Dissimilar words are away.

Vector Algebra Possible

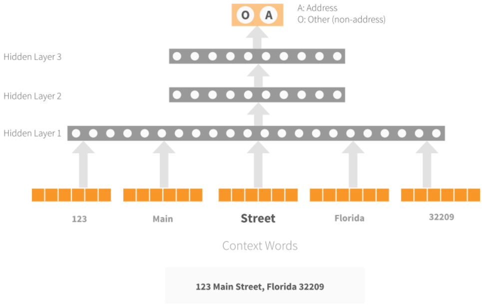
vec[queen] - vec[king] = vec[woman] - vec[man]



(Ref: https://www.tensorflow.org/versions/r0.12/tutorials/word2vec/index.html)

Classifier

Similar words are nearby

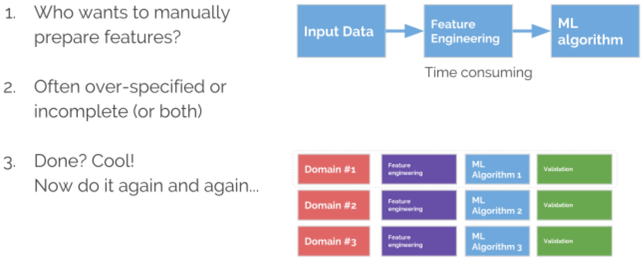


(Ref: Deep Learning for Natural Language Processing - Sihem Romdhani)

Here "Street" is in the context of an Address, thus gets classified that way.

Reasons for Applying Deep Learning to NLP

Automatic Representation Learning



(Ref: A not-so-short introduction to Deep Learning NLP - Francesco Gadaleta, PhD)

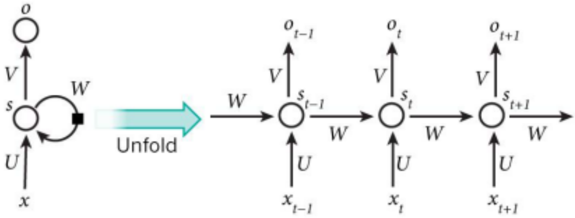
Reasons for Applying Deep Learning to NLP

Learning from unlabeled data

- Typical traditional Machine Leaning based NLP requires labeled training data.
- DL based methods like Skip Gram, CBOW generate word2vec by making labels from unlabeled data.

Reasons for Applying Deep Learning to NLP

Human language is seqeuntial and contextual.



(Ref: A not-so-short introduction to Deep Learning NLP - Francesco Gadaleta, PhD)

RNNs serve the purpose well.

Traditional NLP under threat?

- Deep learning models have taken NLP by storm, achieving superior results across many applications.
- Many DL approaches do not model any linguistic knowledge. They view language as a sequence of strings.
- Is this the end of NLP as a separate discipline?

NLP

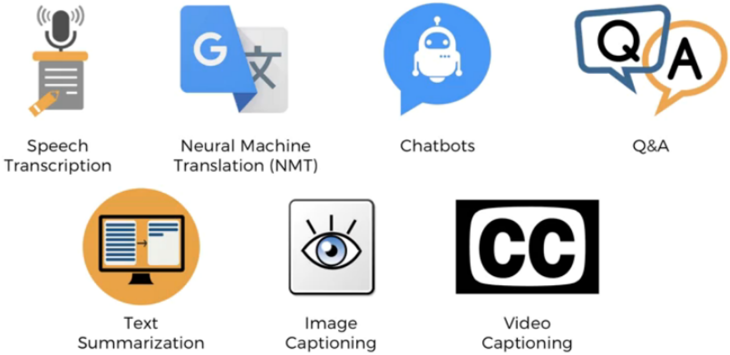
- Rule based systems (since 1960s): Regex
- Machine Learning (since late 1980s): Naive Bayes, SVM, HMM
- Deep Learning (since 2000)

The Promise of Deep NLP

	Get by with rules, search, RegEx, attribute extraction	Welcome to the world of NLP, ML and DL
Social media	Does this social media post contain an offensive word?	Is this social media post offensive?
Legal	Find patents with the terms 'car' and battery', or synonyms	Who is patenting next-gen electrical car batteries?
Support	Find products mentioned in customer emails or phone calls	What is this customer complaining about?
Finance	Extract the fee structure from a mutual fund prospectus	Are UK pensions allowed to invest in this fund?
Healthcare	Extract the patient's blood pressure reading from a note	Does this patient have high blood pressure?

(Ref: Deep Learning for NLU - Dr. David Talby)

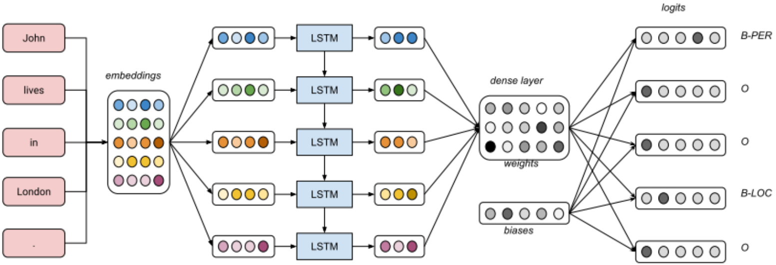
Deep NLP Opportunities



(Ref: Deep Learning and NLP A-Z - Kirill Eremenko)

Deep NLP Algorithms

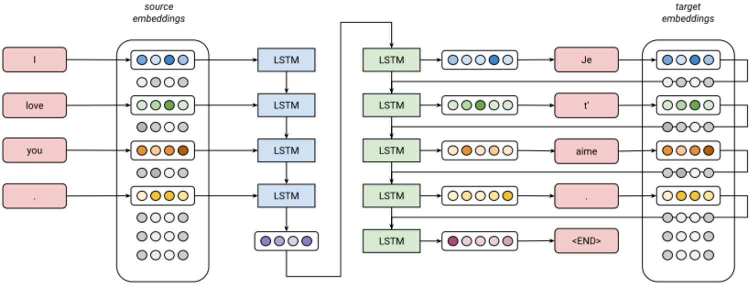
LSTM for sequence labelling



(Ref: Deep Learning for NLP - Yves Peirsman)

Application: named entity recognition

Encoder-Decoder Architecture

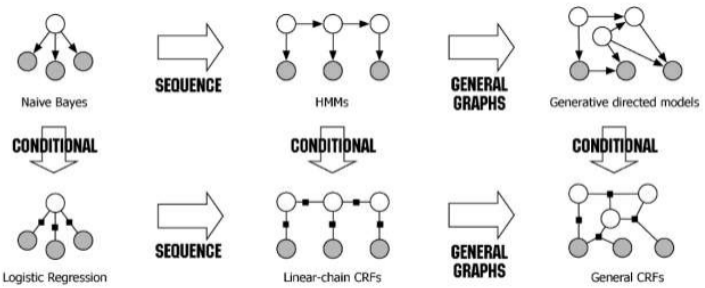


(Ref: Deep Learning for NLP - Yves Peirsman)

Applications: machine translation, text summarization, dialogue modelling, etc

Traditional Named Entity Recognition

The problem



From Sutton & McCallum's [An Introduction to Conditional Random Fields](#).

(Ref: Deep Learning for NLU - Dr. David Talby)

Traditional Named Entity Recognition

Conditional Random Fields (CRFs), "Classic" machine learning approach

81.15%

F-score

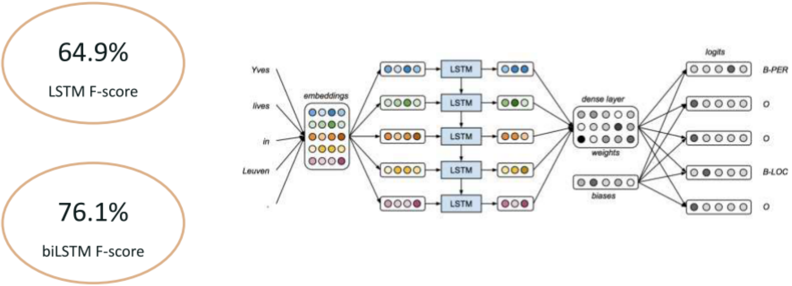
- CoNLL-2003 shared task dataset
- CRF++ Implementation
- Feature engineering:
  - the token itself
  - Its Bigram & trigram
  - Their prefix & suffix
  - Its part of speech
  - Its chunk type
  - Does it start with a capital?
  - Is it uppercase?
  - Is it a digit?
  - Surrounding context words

From Yves Peirsman's Named Entity Recognition and the Road to Deep Learning

(Ref: Deep Learning for NLU - Dr. David Talby)

Deep Named Entity Recognition

LSTM



From Yves Peirsman's [Named Entity Recognition and the Road to Deep Learning]

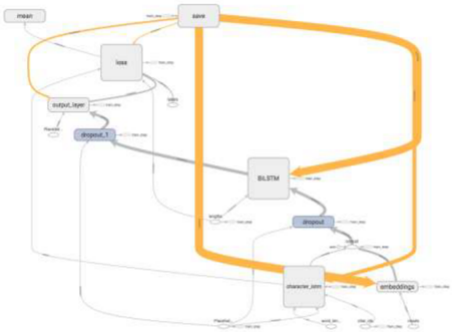
(Ref: Deep Learning for NLU - Dr. David Talby)

Deep Named Entity Recognition

Bi-LSTM

89.3%

F-score



From Yves Peirsman's [Named Entity Recognition and the Road to Deep Learning]

(Ref: Deep Learning for NLU - Dr. David Talby)

Summary of DL algos for NLP

DL Algorithms	NLP Usage
Neural Network (NN) - Feed forward	• POS, NER, Chunking • Entity and Intent Extraction
Recurrent Neural Networks (RNN)	• Language Modeling and Generating Text • Machine Translation • Question Answering System • Image Captioning - Generating Image Descriptions
Recursive Neural Networks	• Parsing Sentences • Sentiment Analysis • Paraphrase Detection • Relation Classification • Object Detection
Convolutional Neural Network (CNN)	• Sentence / Text Classification • Relation Extraction and Classification • Sentiment classification; Spam Detection or Topic Categorization. • Classification of Search Queries • Semantic relation extraction

(Ref: Engineering Intelligent NLP Applications Using Deep Learning – Part 2 Saurabh Kaushik)

Conclusion

- Deep learning has simplified feature engineering in many cases (it certainly hasn't removed it)
- Less feature engineering is leading to more complex machine learning architectures
- Most of the time, these model architectures are as specific to a given task as feature engineering used to be.
- The job of the data scientist will stay sexy for a while (keep your fingers crossed on this one).

What next?

- Coursera : Dr Radev's NLP course (https://www.coursera.org/learn/natural-language-processing)
- Course: Deep NLP By Richard Sochar (Stanford)
- Book: Natural Language Processing with Python

References

- Natural Language Processing - Information Extractoin, Christopher Manning
- Word2Vec - (Girish K.,Sivan Biham & Adam Yaari, Adrian Colyer)
- Deep Learning for Natural Language Processing - Sihem Romdhani
- Notebooks and Material @ https://github.com/rouseguy/DeepLearningNLP_Py

Thanks . . .

- Feel free to follow me at:
  - Github (github.com/yogeshhk) for open-sourced Data Science training material, etc.
  - Kaggle (www.kaggle.com/yogeshkulkarni) for Data Science datasets and notebooks.
  - Medium (yogeshharibhaukulkarni.medium.com) and also my Publications:
    - Desi Stack https://medium.com/desi-stack
    - TL;DR,W,L https://medium.com/tl-dr-w-l
- Office Hours: Saturdays, 2 to 5pm (IST); Free-Open to all; email for appointment.
- Email: yogeshkulkarni at yahoo dot com