Recent advances in Natural Language Processing

Ekaterina Kochmar
What is NLP?

NLP tasks and recent advances

Is “genuine” understanding possible?
What is Natural Language Processing?
Why Natural Language Processing

Because language is inherently very interesting
Why Natural Language Processing

- Because language is inherently very interesting
- Because it’s the primary means of communication
Why Natural Language Processing

- Because language is inherently very interesting
- Because it’s the primary means of communication

- Because it can help you achieve your goals in other domains (“downstream tasks”)
Information Search

When you search for information on the web

- When is next pydata


PyData Cambridge 2019 | 15 - 17 November 2019...
https://cambridgespark.com/pydata-cambridge-2019
From predicting words ...

- Predictive keyboards have become a widely used feature on most phones
- It helps with speeding up typing text and autocorrecting...
to predicting whole conversations

- Google’s Smart Reply can now put together email responses for you
  - Saves time and effort
  - Helps avoid misspellings
  - Often very accurate!

Conversing with machines

Why NLP matters

- Big tech companies use it
- A lot of support in Python community
How is this achieved?
How does NLP help with these tasks?

- A combination of linguistic analysis tools and Machine Learning techniques
Hey Siri, how many cups are in a gallon?

Hey Siri
Call to Siri

how many
This question is about quantities

cups, gallons
The answer needs to connect cups to gallons
Machine Learning in a nutshell

Learning (training) phase

Data

Features

Class

c1

c2

Learn a function \( f \)

Prediction (testing) phase

New data

Features

Apply function \( f \)

https://www.manning.com/books/essential-natural-language-processing
Machine Learning in NLP

- A lot can be achieved by learning patterns from data
- Language is creative but certain things are quite predictable: e.g. how many options are there to continue this phrase?

Would you like to grab some cof_

- We can learn what’s following based on the data for characters, words, and even whole dialogues!
Beyond learning patterns

- A number of things can’t be learned on a pattern basis
- What does it mean for a machine to understand language?
  - Can a machine **understand** and reason about content: e.g., make conclusions, answer questions, show verbal reasoning abilities?
  - Can a machine **generate** free-form text?
Can machines show such abilities?

Deep Learning Machine Beats Humans in IQ Test

Computers have never been good at answering the type of verbal reasoning questions found in IQ tests. Now a deep learning machine unveiled in China is changing that.

CNN article:

Document

The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the “Top Gear” host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon “to an unprovoked physical and verbal attack.” . . .

Query

Producer X will not press charges against Jeremy Clarkson, his lawyer says.

Answer

Oisin Tymon
Has a human or a machine written this?

1 “A shallow magnitude 4.7 earthquake was reported Monday morning five miles from Westwood, California, according to the U.S. Geological Survey. The temblor occurred at 6:25 a.m. Pacific time at a depth of 5.0 miles.”

2 “Apple’s holiday earnings for 2014 were record shattering. The company earned an $18 billion profit on $74.6 billion in revenue. That profit was more than any company had ever earned in history.”
1. “A shallow magnitude 4.7 earthquake was reported Monday morning five miles from Westwood, California, according to the U.S. Geological Survey. The temblor occurred at 6:25 a.m. Pacific time at a depth of 5.0 miles.”

2. “Apple’s holiday earnings for 2014 were record shattering. The company earned an $18 billion profit on $74.6 billion in revenue. That profit was more than any company had ever earned in history.”
How do machines “understand” language?

---

Mikolov et al. (2013) showed that a machine can reason like a human, e.g. it can solve analogy tasks

Man is to woman as king is to _______?

We may be thinking along the following lines:

meaning(king) - meaning(man) + meaning(woman) = what we are looking for

Mikolov et al. (2013). Linguistic Regularities in Continuous Space Word Representations
How do machines “understand” language?

- To machines, text is a sequence of symbols
- What machines are good with is numbers
- Let’s build a numerical representation of a word that captures its meaning
- Imagine we can represent word meaning using a vector
- Then we can apply all sorts of operations:
  - Interpret similarity as distance in vector space
  - Find the word we are looking for using simple operations on vectors
Analogy task solved with vectors

Here, we equate

\[
\text{meaning}(\text{king}) - \text{meaning}(\text{man}) + \text{meaning}(\text{woman}) = \text{meaning}(\text{queen})
\]

to operations on vectors

\[
\text{vector}(\text{king}) - \text{vector}(\text{man}) + \text{vector}(\text{woman}) = \text{vector}(\text{queen})
\]
How do we learn such a vector representation?

Earlier approaches relied on **distributional semantics**

> You shall know a word by the company it keeps

(Firth, J. R. 1957)

- queen
  - royal
  - gives speeches
  - crown
  - female

- king
  - royal
  - gives speeches
  - crown
  - male
Learn the properties from context

Her Majesty the Queen
The Queen's speech during the State Visit to...
Buckingham Palace is the Queen's official London residence...
The Crown of Queen Elizabeth
The Queen Mother

<table>
<thead>
<tr>
<th></th>
<th>she</th>
<th>he</th>
<th>crown</th>
<th>palace</th>
</tr>
</thead>
<tbody>
<tr>
<td>queen</td>
<td>55</td>
<td>2</td>
<td>32</td>
<td>29</td>
</tr>
</tbody>
</table>
Build vectors based on these context properties

Note that this is extended to a multi-dimensional space, for as many dimensions as there are properties that define word meaning.
High quality pre-trained word vectors

- **Word embeddings** - word vectors trained on large amounts of data such that:
  - Distance between vectors representing *similar* meaning is **minimised**
  - Distance between vectors representing *dissimilar* meaning **maximised**

- **Word2vec** (Mikolov et al., 2013)
  - [https://code.google.com/archive/p/word2vec/](https://code.google.com/archive/p/word2vec/)

- **GloVe** (Pennington et al., 2014)
  - [https://nlp.stanford.edu/projects/glove/](https://nlp.stanford.edu/projects/glove/)
Further analogy tasks

The analogy tasks have been extended to a number of other scenarios:

- Famous personalities - field (Einstein - scientist)
- Country - food (Japan - sushi)
- Country - capital (Russia - Moscow)
Easy to implement with python and spaCy

```python
import spacy

text = u'Amsterdam Ankara Athens Australia Barcelona Beijing Berlin Brazil China France Germany Greece Italy Japan Lisbon London Madrid Moscow Paris Poland Portugal Rome Russia Spain Switzerland Tokyo Turkey Venice Warsaw'

nlp = spacy.load('en')

words = nlp(text)

for word in words:
    print(word)
    print (word.vector[:5])
```

- **nlp** - spaCy’s NLP pipeline that gives you access to word vectors

Easy to implement with python and spaCy

```python
def analogy_task(country):
    question = "Russia is to Moscow as " + country
    text = nlp(question)
    source1 = text[0]
    source2 = text[3]
    target1 = text[5]

    max_sim = 0.0
    target2 = "N/A"

    target2_vector = source2.vector - source1.vector + target1.vector

    for word in words:
        if not (str(word) == str(target1) or str(word) == str(source1) or str(word) == str(source2)):
            current_sim = cosine(target2_vector, word.vector)
            if current_sim >= max_sim:
                max_sim = current_sim
                target2 = word

    print(question)
    print("It is to " + str(target2))

countries = ["China", "France", "Germany", "Greece", "Italy", "Japan", "Poland", "Portugal", "Spain", "Turkey"]

for country in countries:
    analogy_task(country)
```

Further advances in language understanding
Machines learn language better by using a deep understanding of words

Devin Coldewey  @techcrunch  /  6:53 pm BST  •  June 15, 2018

Challenges for word vector representations

- Word embeddings incorporate previous knowledge at the first step (first layer of the network) only → “shallow”
- Challenges:
  - Similar words with opposite meaning often used in similar contexts, so may result in very similar vectors (e.g., “hot” and “cold”)
  - One still needs to derive the meaning of sentences and texts from word vector building blocks
  - Ambiguous words get single representation

https://ruder.io/nlp-imagenet/
Can one do better?

- From “shallow” models to deep pre-trained language models.
- Rather than initialising the first layer only, train the whole model hierarchically.
- This move has been called the “ImageNet moment for NLP”.
- That means having a very powerful pre-trained model that captures peculiarities of meaning and can be easily fine-tuned to any new data (called transfer learning).

https://ruder.io/nlp-imagenet/
Can one do better?

Learning features that are likely to generalise beyond a particular dataset to new tasks in the problem domain

https://distill.pub/2017/feature-visualization/
How to train a better model?

- Take a **sufficiently large dataset** (in order of millions of training examples)
- Define a task that is **representative of the problem space**
- **Language modelling** – predicting what comes next – ticks all the boxes

The service was poor, **but** the food was ___

https://ruder.io/nlp-imagenet/
Training with a language model objective

Royal visit by the **Queen** and ___

I moved my **queen** to ___
Example: ULMFiT

- Pre-train with a language model objective
- Fine-tune on any task of your choice

Pre-trained models

- Universal Language Model Fine-Tuning (ULMFiT)
  - [http://nlp.fast.ai/category/classification.html](http://nlp.fast.ai/category/classification.html)
- ELMo deep contextualised word representations
  - [https://allennlp.org/elmo](https://allennlp.org/elmo)
- OpenAI Transformer model
  - [https://github.com/openai/finetune-transformer-lm](https://github.com/openai/finetune-transformer-lm)
- Deep Bidirectional Transformers (BERT)
  - [https://github.com/google-research/bert](https://github.com/google-research/bert)
What does it mean for a machine to “understand”? 
The Great A.I. Awakening

How Google used artificial intelligence to transform Google Translate, one of its more popular services — and how machine learning is poised to reinvent computing itself.

BY GIDEON LEWIS-KRAUS  DEC. 14, 2016

https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html?_r=0
Chinese Room Thought Experiment

- - -

- Proposed by John Searle (1980)
- It is perfectly possible to construct a machine that learns to take the correct sequence of actions to solve the task without "genuine" understanding

source: wikicommons
What does it mean to genuinely understand for human?

- Most human beings understand what “water” means and will be able to describe most of its properties.
- Unfortunately, many don’t understand that water is electrical conductor.
- Does this constitute genuine understanding of the concept of “water”?

https://medium.com/@tdietterich/what-does-it-mean-for-a-machine-to-understand-555485f3ad40
What does it mean to genuinely understand for a machine?

---

Who did IBM’s Deep Blue system defeat?

About 7,470,000 results (0.75 seconds)

Kasparov

The victor was even more unusual: IBM supercomputer, Deep Blue. In defeating Kasparov on May 11 1997, Deep Blue made history as the first computer to beat a world champion in a six-game match under standard time controls. 11 May 2017

Twenty years on from Deep Blue vs Kasparov: how a chess ...
theconversation.com › twenty-years-on-from-deep-blue-vs-kasparov-how-a…

when

adverb
at what time.
"when did you last see him?"

adverb
at or on which (referring to a time or circumstance).
"Saturday is the day when I get my hair done"

conjunction
1. at or during the time that.
"I loved maths when I was at school"

2. after which; and just then (implying suddenness).
"he had just drifted off to sleep when the phone rang"

https://medium.com/@tdietterich/what-does-it-mean-for-a-machine-to-understand-555485f3ad40
Conclusion

- The notion of “genuine” understanding and intelligence is a fuzzy one
- Rather than asking for comprehensive “genuine” understanding define specific tasks the machine can excel at
- It is important to claim credit for successes in the field, as well as take ownership for shortcomings

https://medium.com/@tdietterich/what-does-it-mean-for-a-machine-to-understand-555485f3ad40
If you would like to learn more about NLP

- A friendly practical introduction to NLP
- Covers a range of topics from information search to sentiment analysis to summarisation
- Use ctwpycbg19 to get a 40% discount
- Free e-book copies available (first 5 requests) - let me know

https://www.manning.com/books/essential-natural-language-processing
Thank you!