

Natural Language Processing

COMP2002

Lauren Ansell

Today's topics:

- Q&A systems
- Deep learning for NLP

Session learning outcomes - by the end of today's lecture you will be able to:

- Explain the challenges involved in building Q&A systems
- Explain the role of sentiment analysis in NLP tools
- Understand how deep learning is used within NLP

“Alexa – tell me about the University of Plymouth”

A longstanding wish has been AI that can respond to verbal questions.

“Natural language” follows grammatical rules but not so strictly that it can be understood by a computer.

They cannot understand context.

Sentiment and sarcasm are lost.

Who are “they”?

Lack of common sense.

ELIZA – An Early NLP Tool

Script-based NLP tool from the 1960s.

Pattern matching is used to respond to users and make them feel as though the responder is interested in them.

Conversations don't take long to get strange. . .

Talk to Eliza by typing your questions and answers in the input box.

```
> Hello, I am Eliza. I'll be your therapist today.  
* Hi Eliza. I'm feeling a little sad today.  
> Do you believe it is normal to be feeling a little sad today?  
* I don't think so.  
> Don't you really think so?  
* No I don't  
> Do you wish to be able to ?  
* I'd prefer not to be!  
> I see. Let's try another topic and we will come back to that issue later.
```

TYPE HERE

Talk to Eliza

```
> Hello, I am Eliza.  
* Hi Eliza  
> How are you today.. What would you like to discuss?  
* I'm feeling very sad because my cat is missing.  
> How long have you been feeling very sad because your  
cat is missing?  
* Since it went missing yesterday.  
> Say, do you have any psychological problems?  
* Well I'm sad...  
> How long have you been sad?
```

Input:

IBM Watson – Q&A In The 2010s

In 2011 IBM entered IBM Watson in a Jeopardy! contest against two of the game's champions.

Watson using machine learning and massive repositories of data (books, Wikipedia, journals...) to “learn”.

Watson answers a question by finding a range of answers that are ranked to identify the most likely result trained on a large dataset of past questions and answers.

“An account of the Principalities of Wallachia and Moldavia inspired this author's most famous novel”.

A: Who is Bram Stoker?

Is this intelligence?

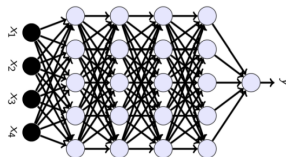
Speech Recognition

State-of-the-art speech recognition has been done with deep neural networks since 2012.

Offered a substantial improvement without requiring an understanding of knowledge.

Reduced performance with background noise, odd accents, unusual phrases. . .

Performance is extremely good but not human level – is language understanding required?



**This is what
I'm saying...**

Sentiment Analysis

“That’s a very brave idea”

Could mean that the speaker is praising the idea or is putting it down.

We want to identify the true sentiment behind a statement automatically – hard for a computer to do – humans we use context and background knowledge.

Large amounts of data to be mined in customer reviews and social media.

Multiple applications.

Lexicons

A catalogue of a language's words.

Lexicon-based sentiment analysis uses lists of words and phrases that are linked to different emotions to label the words.

Examples:

- Bing
- Afinn
- NRC
- Loughran

Example: "A chill ran up Lorkin's spine, and his heart missed a beat. Father made this! He turned it over and over, the stone catching the light. Did Father know stone-making? Surely not. The answer was suddenly clear to him. It must be a blood gem. The implications of that hit him like a slap. 'You were in communication with him all along!' "

Bing Lexicon

Developed by Minqing Hu and Bing Liu starting with their 2004 paper.

It contains a total of 6,786 word - 2,005 positive and 4,781 negative.

In our example text, we obtain the following results:

Word	Sentiment
chill	negative
missed	negative
clear	positive
gem	positive
like	positive
slap	negative

To obtain a numerical score, we assign the value 1 to positive words and -1 to negative.

Afinn Lexicon

Developed by Finn Årup Nielsen in 2009.

It consists of 2,477 words with 878 positive and 1,598 negative.

The lexicon assigns words a score between -5 and +5.

In our example text, we obtain the following results:

Word	Value
like	2
clear	1
missed	-2

The original lexicon contains some multi-word phrases, but they are excluded here.

NRC Emotion Lexicon

A list of 5,636 English words which links words to eight emotions and two sentiments.

The emotions are: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust.

In our example text, we obtain the following results:

Word	Sentiment	Word	Sentiment
spine	anger	slap	negative
stone	anger	spine	positive
hit	anger	gem	positive
slap	anger	suddenly	surprise
gem	joy	slap	surprise
spine	negative	father	trust
stone	negative	communication	trust
hit	negative		

Loughran Lexicon

The Loughran lexicon contains 2709 words.

The lexicon is widely used in finance for financial statement.

Similar to the NRC lexicon, words are assigned to groups which are: constraining, litigious, negative, positive, superfluous, and uncertainty.

In our example text, we obtain the following results:

Word	Sentiment
missed	negative
suddenly	uncertainty

Document Classification

The aim is to predict the attributes of documents.

Most simple method to achieve this is *bag-of-words*.

The *bag-of-words* creates a summary of the document based on the words present.

We can include the context of the document by:

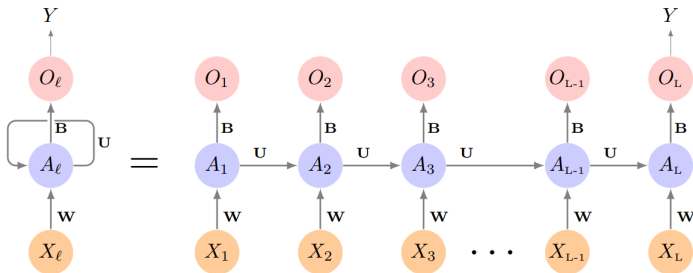
- using more words: *bag-of-n-grams*.
- treating the document as a sequence.

Recurrent Neural Networks

RNNs allow previous outputs to be used as inputs so that history can be incorporated.

Read left to right – in the way you would read a book

Activation passed along to the next timestep to include memory of past inputs



One-hot Encoding

1	2	...	417	418	419	...	998	999	1000
elephant	cheese		bus	excellent	website		apple	video	network
0	0		0	1	0		0	0	0

Given a vocabulary of N words network input has N features.

One of them is set to 1 (the current word).

What if the sentence said “awesome” or “great”?

RNN Example - Sentiment from IMDB Reviews

The training data consists of the 10,000 most frequent words from a corpus of 25,000 reviews.

The aim is to predict the sentiment expressed in the review from the words contained in the review.

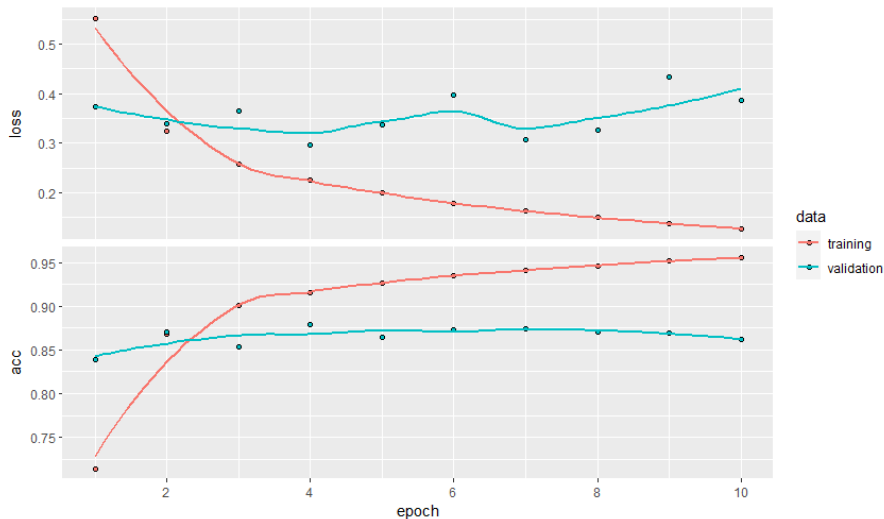
When preparing the data for the RNN, we need to restrict the document sequences to all have the same length, in this case 500 words.

The first layer of the RNN is an embedding layer of size 32.

The second layer is an LSTM with 32 units.

The output layer is a single sigmoid for the binary classification task.

RNN Example - Results



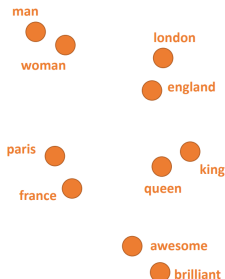
Semantic Spaces

“You shall know a word by the company it keeps” – John R. Firth, linguist

Distributional semantics – e.g., the word “bus”, “car” and “taxi” tend to occur in similar contexts

Requires a large number of dimensions to disaggregate words with multiple meanings.

Hard to visualize – remember we need two dimensional spaces



word2vec

Approach proposed by Google to embed words into a semantic space

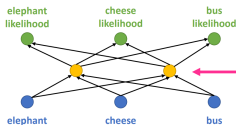
One-hot encoding mapping each word to likely following word

Pairs of words from Google News service

(“approach”, “proposed”), (“proposed”, “Google”), (“Google”, “embed”),
 (“embed”, “words”)...

Deep network with 300 hidden neurons mapping N inputs to N outputs (N words)

Hidden neurons' activation for a given word is the word vector



Hidden layer's activation
for a word is its word
vector – used as input to
NLP models to capture
semantic space

Language Translation

Can be done live – 2012 speech from Microsoft was translated into Chinese on the fly.

Google Translate offers translation services – small amounts of text are OK, larger blocks can show oddities.

Social media translates your posts

Bengali - English translation

“As one ball appealed against him, he burned in oil and eggplant and hit down the wicket”

“In one delivery, someone appealed against them, and they lost their marbles and went down the track to hit the ball with all their might in the next delivery.”

Captioning – Translating Images To Phrases

What am I looking at?

As far as the computer is concerned this is much the same task – translate between pixel values and text.

Use a convnet to work out what's in the image – rather than outputting a class, use this as the input to a translation tool

CaptionBot (from Microsoft)



I think it's a road next to a body of water.



I think it's a large ship docked.

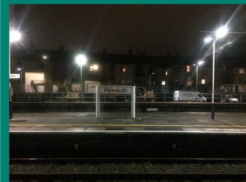


©David Walker

How did I do?



I think it's a sign on the side of a road.



I think it's a black box with a blue and green design.



How did I do?



The good

The less good...



Introduction To Large Language Models

Definition: Large language models are machine learning models that are trained on a large dataset of text and are capable of generating human-like text. They can be used for a variety of tasks, including language translation, summarization, and answering questions.

Examples: Some examples of large language models include GPT-3 (OpenAI), Transformer (Google), and BERT (Google).

How Large Language Models Work

Training: Large language models are typically trained using supervised learning, where the model is fed a large dataset of text and a corresponding label (e.g., the correct translation of a sentence in another language). The model then learns to predict the label based on the input text.

Architecture: Large language models typically use a transformer architecture, which consists of multiple layers of attention mechanisms that allow the model to process input text and make predictions.

Pre-training: Many large language models are pre-trained on a large dataset and then fine-tuned on a smaller, specific task-specific dataset. This allows the model to learn general language patterns and then specialize in a particular task.

Applications And Limitations

Applications: Large language models have a wide range of applications, including language translation, summarization, question answering, and text generation. They have been used in industries such as healthcare, finance, and customer service.

Limitations: Despite their impressive capabilities, large language models have some limitations. They can be expensive to train and require a large amount of computational resources. They can also produce biased or offensive output if the training data contains biased language. Additionally, they may not be able to generalize to tasks or situations that are not represented in the training data.

Another Application Of LLMs

Producing lecture slides

The text for the last three slides was entirely generated by ChatGPT (<https://openai.com/blog/chatgpt/>)

The only human interaction was copying & pasting into PowerPoint and highlighting some key words

Time taken – approximately 5 minutes

Describe large language models in enough detail for three PowerPoint slides



Sure! Here is some information that could be included in three PowerPoint slides about large language models:

Slide 1: Introduction to Large Language Models

- Definition: Large language models are machine learning models that are trained on a large dataset of text and are capable of generating human-like text. They can be used for a variety of tasks, including language translation, summarization, and answering questions.
- Examples: Some examples of large language models include GPT-3 (OpenAI), Transformer (Google), and BERT (Google).

Slide 2: How Large Language Models Work

- Training: Large language models are trained using supervised learning, where the model is fed a large dataset

Regenerate response

ing label (e.g., the correct

Natural Language Processing

- The task of getting computers to process language in the a human-like fashion
- Complicated – humans use context and sentiment to express themselves
- Deep learning provides tools that are helpful
- Some successes – e.g. speech recognition
- More work to be done in other areas (e.g. sentiment analysis, translation)