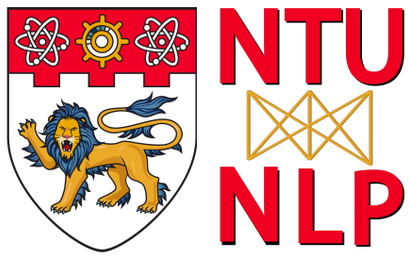
Deep Learning for Natural Language Processing

CE/CZ 4045

Shafiq Joty



Lecture 1: Introduction

Lecture Plan

- Logistics
- What is NLP (a quick recap)
- What is Deep Learning
- Why deep learning for NLP
- Outline
- Knowing the target group

About me

Home

Papers

Research

Teaching

Students

CV

Shafiq Rayhan Joty

Assistant Professor
NTU Natural Language Processing Group
Office: Block N4, 02c-79
School of Computer Science and Engineering
Nanyang Technological University, Singapore

Email
Phone
Google Scholar
Github

Senior Manager (Joint appointment)
Salesforce AI Research
Office: Floor 40, Suntec City Tower 2

Email
Phone

Short Biography

- Assistant Professor, Nanyang Technological University (NTU), Singapore [Jul'17 -]
- Senior Manager, Salesforce AI Research, Singapore [Feb'19 -]
- Research Scientist, Qatar Computing Research Institute (QCRI), Doha [Jan'14 - Jul'17]
- PhD in Computer Science, University of British Columbia (UBC), Vancouver [Sep'08 - Dec'13]

Research Interests

Natural Language Processing

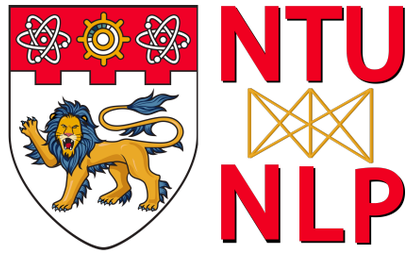
- NLP tools (Syntax, Semantics and Discourse)
- Multi-lingual NLP (Machine Translation, Cross-lingual tasks)
- NLP Applications (QA, Summarization, Dialogue)
- Robust/adversarial NLP
- Multi-modal NLP (Image/Video Captioning)

Machine Learning

- Deep Learning
- Probabilistic Graphical Models
- Reinforcement Learning

https://raihanjoty.github.io/

Course Logistics



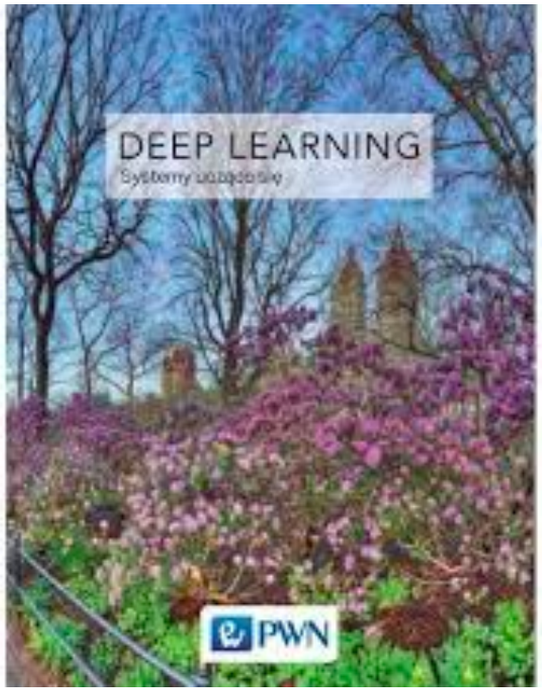
- Time:
  - Lecture: Wednesday 9:30 am –11:30 am, Online (Teams)
  - Tutorial: Friday 1:30 pm - 2:30 pm, Online (Teams)
  - Part-Time: Friday: 6:30 pm - 9:30 (Lecture + tutorial)

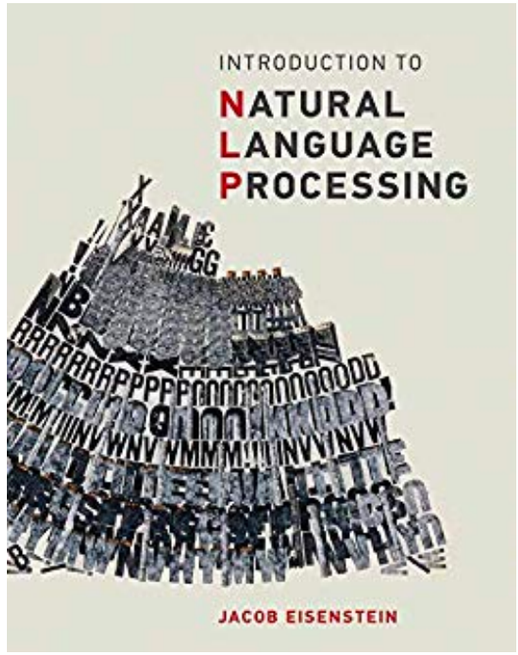
- Course materials
  - Slides/Materials will be uploaded before each lecture
- Slido: **www.slido.com** using the event code **#94155**

Optional Textbooks

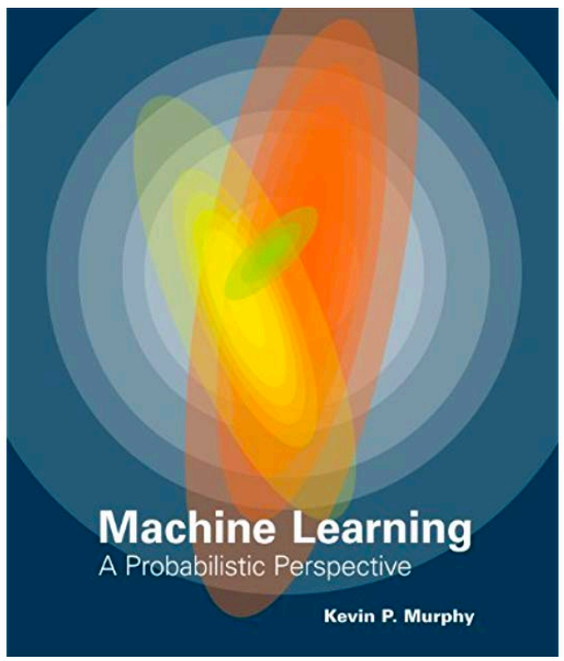
NTU NLP

- Deep Learning by Goodfellow, Bengio, and Courville free online



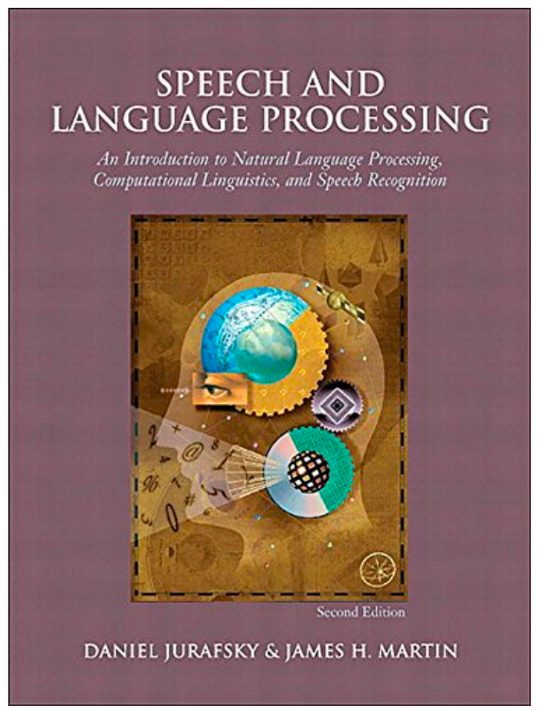


- Machine Learning — A Probabilistic Perspective by Kevin Murphy

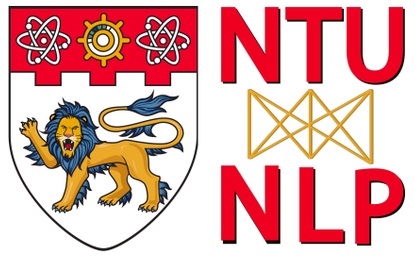


- Natural Language Processing by Jacob Eisenstein free online

- Speech and Language Processing by Dan Jurafsky and James H. Martin (3rd ed. draft)

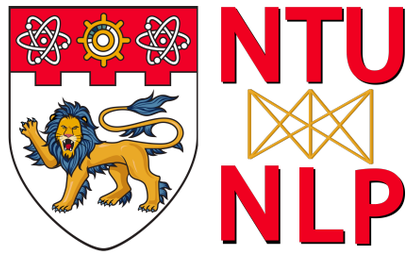


Prerequisites



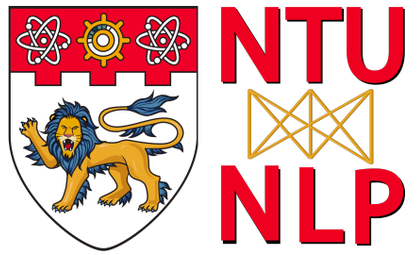
- Proficiency in Python
  - See Supplementary material (Course website)
- Multivariate Calculus, Linear Algebra
- Basic Probability and Statistics
- Fundamentals of Machine Learning
  - Loss functions, Taking simple derivatives, Performing optimization with gradient descent

Course Objectives



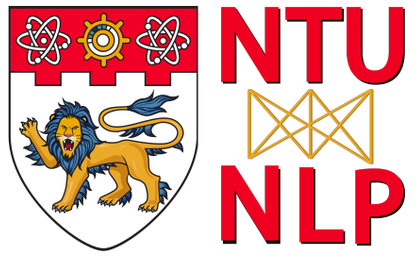
- Understanding of the deep learning methods in NLP
- A big picture understanding of natural languages and the difficulties in understanding and generating them
- An understanding of and ability to build systems for some of the major problems in NLP
  - Word meaning, parsing, named entities, machine translation, question answering, dialogue systems, cross-lingual applications

Approach



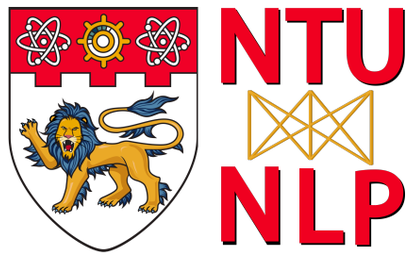
- **Thorough and Detailed:** How to write from scratch, debug and train deep neural models
- **State of the art:** Most lecture materials are new (past 1-5 years).
- **Practical:** Focus on practical techniques for training the models
- **Fun:** Cover exciting new advancements in NLP (e.g., Transformer, BERT).
- **Research:** Some materials are advanced (good for research!)

Grading policy



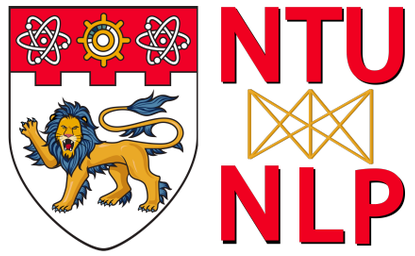
- 1 Quiz: 30%
  - Will be on 13-th week (Friday, Tutorial time)
- 1 Assignment (4-5 members group): 20%
  - Will be released on the 10th week
  - Due end of week-12

Lecture Plan



- Course logistics
- What is NLP?
- What is Deep Learning
- Why deep learning for NLP
- Knowing the target group

What is NLP

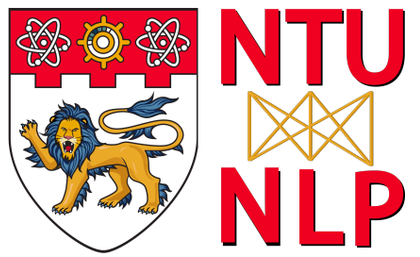


We study formalisms, models and algorithms to allow computers to perform useful tasks involving knowledge about human languages.

Useful Tasks

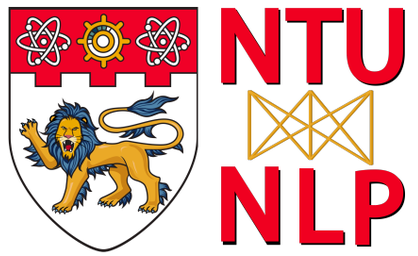
- **Conversational agents:**
  - AT&T "How may I help you?" technology
  - Apple SIRI, Amazon's alexa, Microsoft's Cortana.
- **Summarization:** "Please summarize my discussion with Sue about NLP" "What people say about the new Nikon 5000?"
- **Machine Translation:** Google translate (75B USD$ industry)
- **Text Generation:** Data2Text, Table2Text, Video/Image2Text
  - ARRIA world leader in NLG- when it floated on London's Alternative Investment Market (AIM) in 2013, it was valued at over £160 million

Useful Tasks



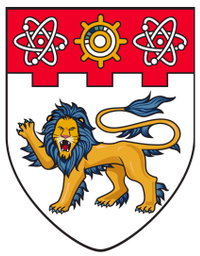
- **Question answering:** "Was 1991 an El Nino year? ....Was it the first one after 1982?" "Why was it so intense?"
  - IBM Watson Jeopardy (now medicine)
- **Document Classification:** spam detection, news filtering
- **Information Extraction:** Knowledge graphs (in Search)
- **Speech:**
  - speech recognition, text to speech synthesis.
- **Multimodal:** Image-to-text (image/video captioning, VQA)

What is NLP



We study formalisms, models and algorithms to allow computers to perform useful tasks involving knowledge about human languages.

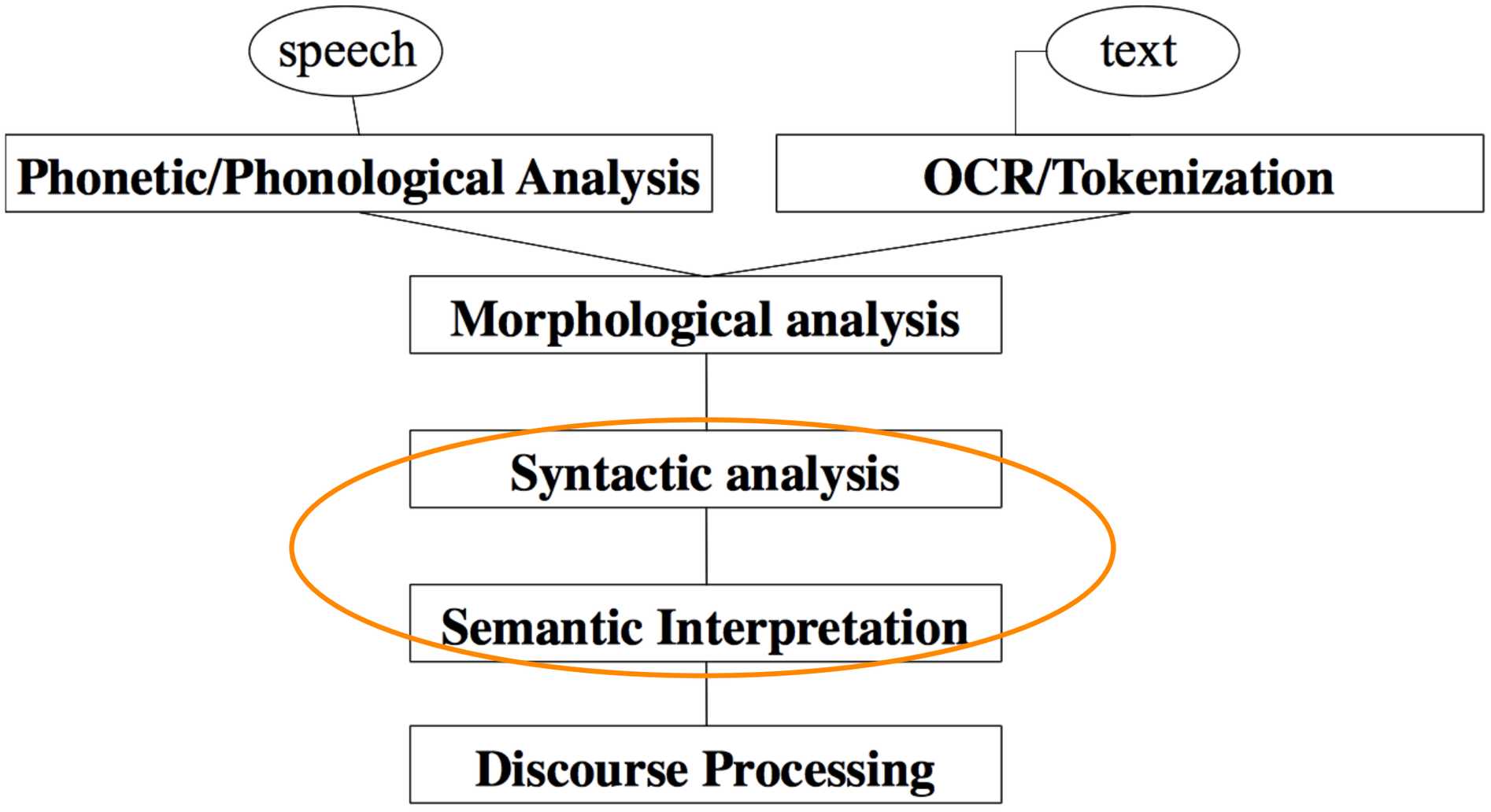
Knowledge about Language



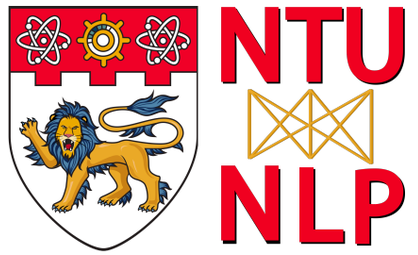
NTU NLP

- Phonetics and Phonology (sounds)
- Morphology (structure of words)
- Syntax (structure of sentences)
- Semantics (meaning)
- Pragmatics (language use)
- Discourse and Dialogue (units larger than single utterance)

Knowledge about Language



Natural Language as Signal



- Language is mostly a **discrete/symbolic/categorical** signalling system
  - Dog, Cat, Pen
- The categorical symbols of a language can be encoded as a signal for communication in several ways:
  - Sound, Writing, Gesture
- However, human brain's encoding appears to be a **continuous pattern of activation**, and the symbols are transmitted via **continuous signals** of sound/vision

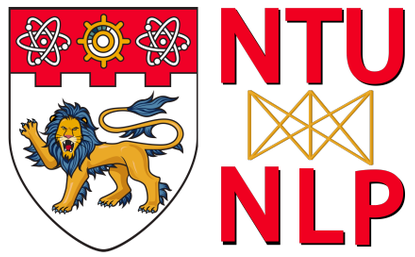
What is NLP

NTU NLP

We study formalisms, models and algorithms to allow computers to perform useful tasks involving knowledge about human languages.

Different Neural Models (This half)
Traditionally (First half):
Finite State Automata, Markov Models, Graphical Models

Lecture Plan



- Course logistics
- What is NLP?
- What is Deep Learning
- Why deep learning for NLP
- Knowing the target group

What is Deep Learning?

- **Deep learning** is a subfield of **machine learning.**
- Most machine learning methods work well because of **human-designed representations/input features.**
- Machine learning becomes just optimizing feature weights to make a good prediction.

Feature	**NER**
Current Word	✓
Previous Word	✓
Next Word	✓
Current Word Character n-gram	all
Current POS Tag	✓
Surrounding POS Tag Sequence	✓
Current Word Shape	✓
Surrounding Word Shape Sequence	✓
Presence of Word in Left Window	size 4
Presence of Word in Right Window	size 4

Features for NER (Finkel et al., 2010)

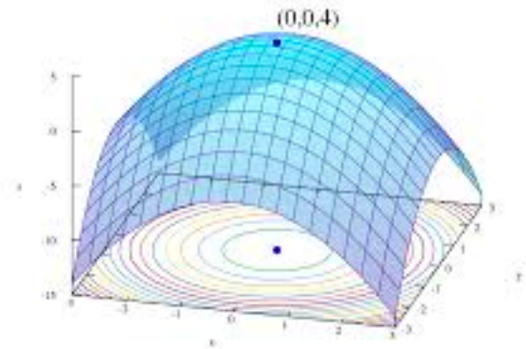
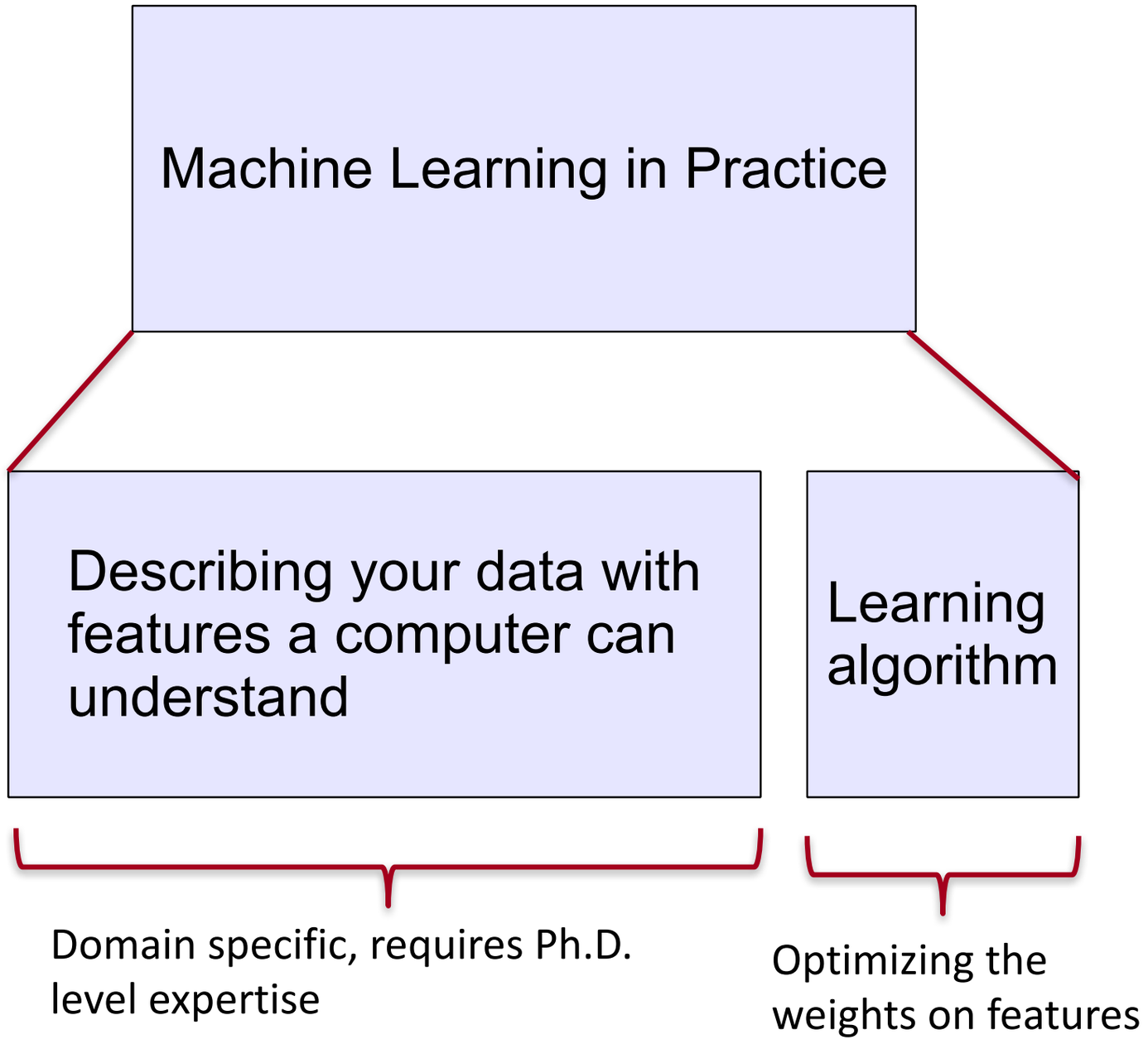


Figure 1: An example of NER application on an example text

Traditional Machine Learning



Source: Stanford 224n

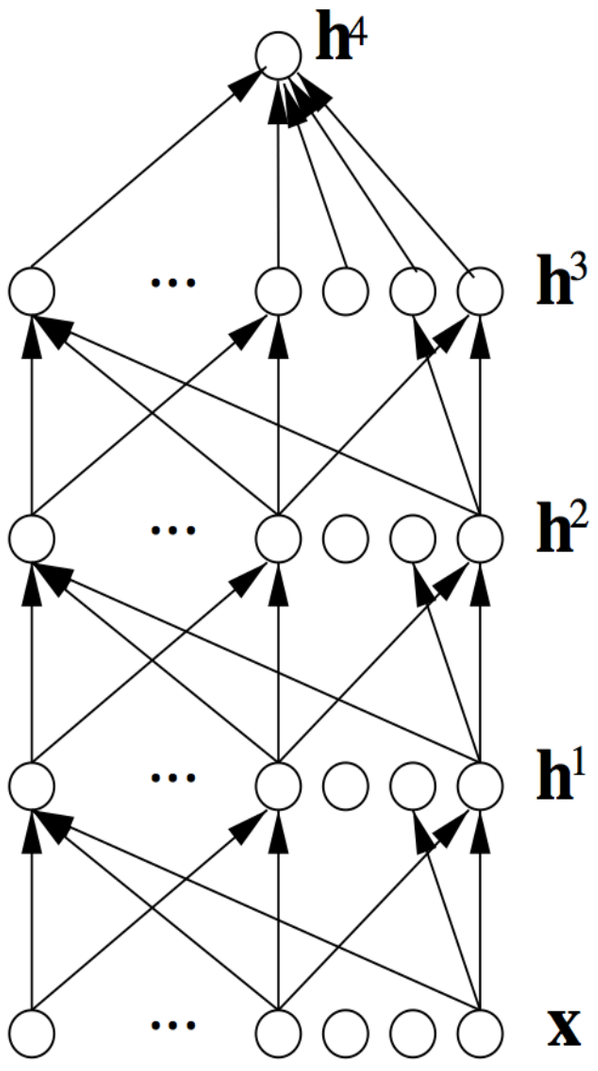
Deep Learning

NTU NLP

- Representation learning attempts to automatically learn good features or representations

- Deep learning algorithms attempt to learn (multiple levels of) representations (here: h^1,h^2,h^3) and an output (h^4)

- From "raw" inputs **x** (e.g. sound, pixels, characters, or words)



Source: Stanford 224n

Why Deep Learning?

NTU NLP

- Manually designed features are often over-specified, incomplete and take a long time to design and validate
- **Learned Features** are easy to adapt, fast to learn
- Deep learning provides a very flexible, learnable framework for **representing** world, visual and linguistic information.
- Deep learning can learn in **unsupervised** (from raw text) and **supervised** (with specific labels like positive/negative) settings

Source: Stanford 224n

Why Deep Learning?

NTU NLP

- In ~2010 **deep** learning techniques started outperforming other machine learning techniques. Why this decade?
  - Large amounts of training data favor deep learning
  - Faster machines and multicore CPU/GPUs favor deep learning
- New models, algorithms, ideas
  - Better, more flexible learning of intermediate representations
  - Effective end-to-end joint system learning
  - Effective learning methods for using contexts and transferring between tasks
  - Better regularization and optimization methods
- **Improved performance** (first in speech and vision, then NLP)

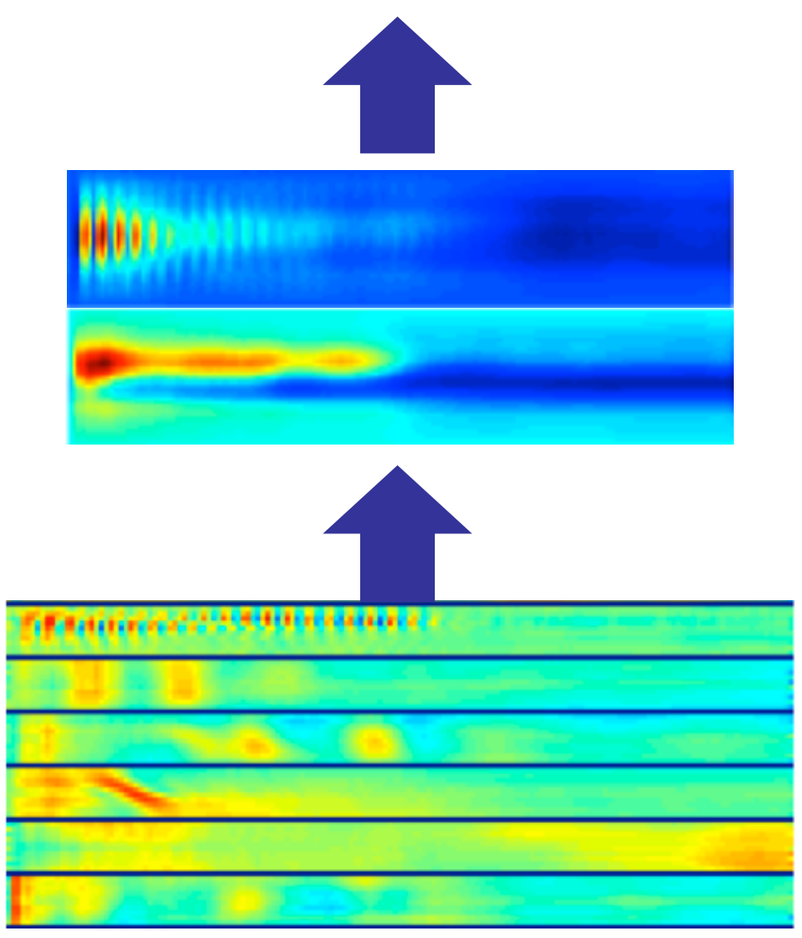
Source: Stanford 224n

Deep Learning in Speech

NTU NLP

- The first breakthrough results of "deep learning" on large datasets happened in speech recognition
- Context-Dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition Dahl et al. (2010)

Acoustic model and WER	RT03S FSH	Hub5 SWB
Traditional features	**27.4**	**23.6**
Deep Learning	**18.5** (–33%)	**16.1** (–32%)



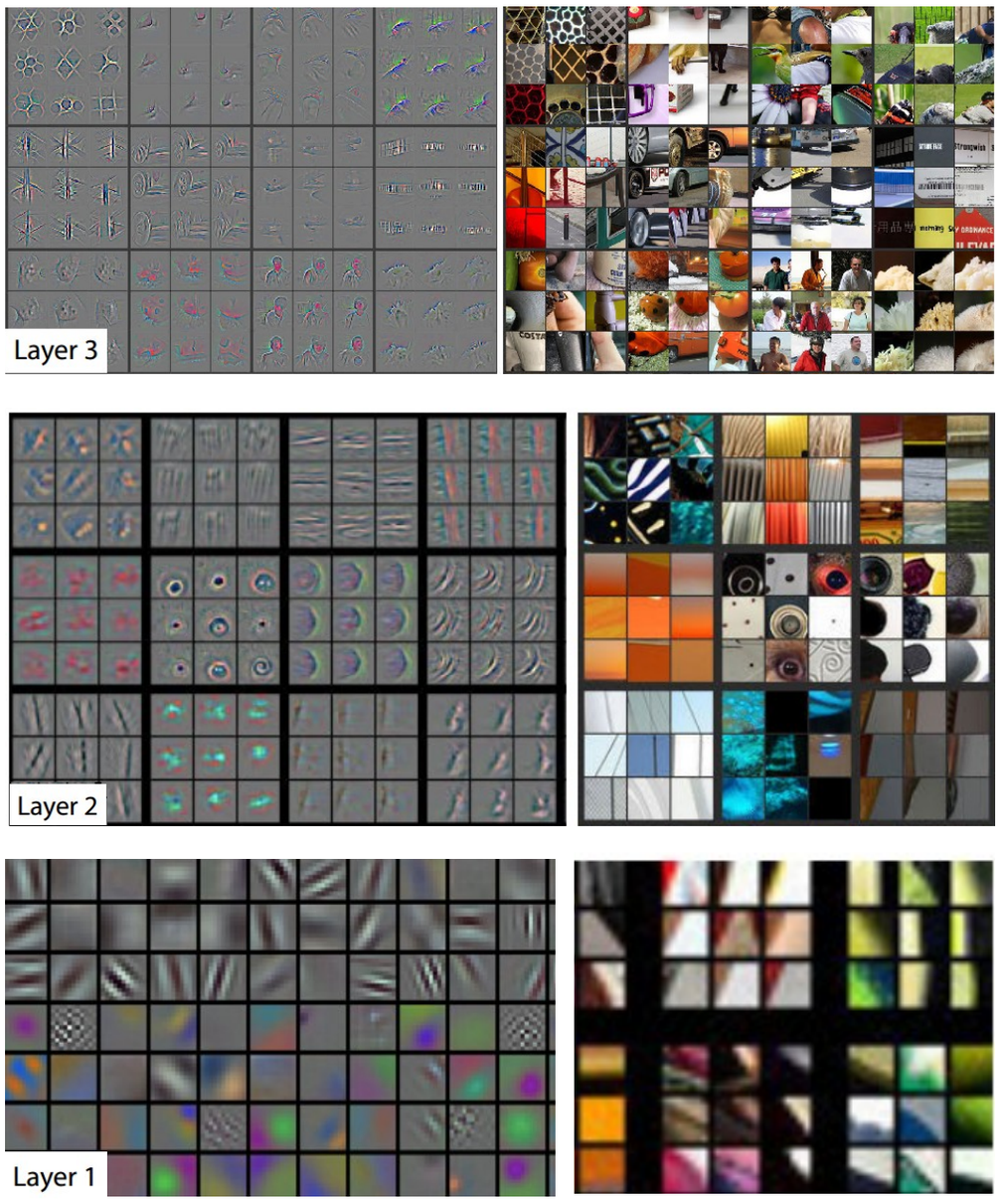
Source: Stanford 224n

Deep Learning in Vision

NTU NLP

First major focus of deep learning groups was computer vision

The breakthrough DL paper:
ImageNet Classification with Deep Convolutional Neural Networks by Krizhevsky, Sutskever, & Hinton, 2012, U. Toronto. 37% error red.



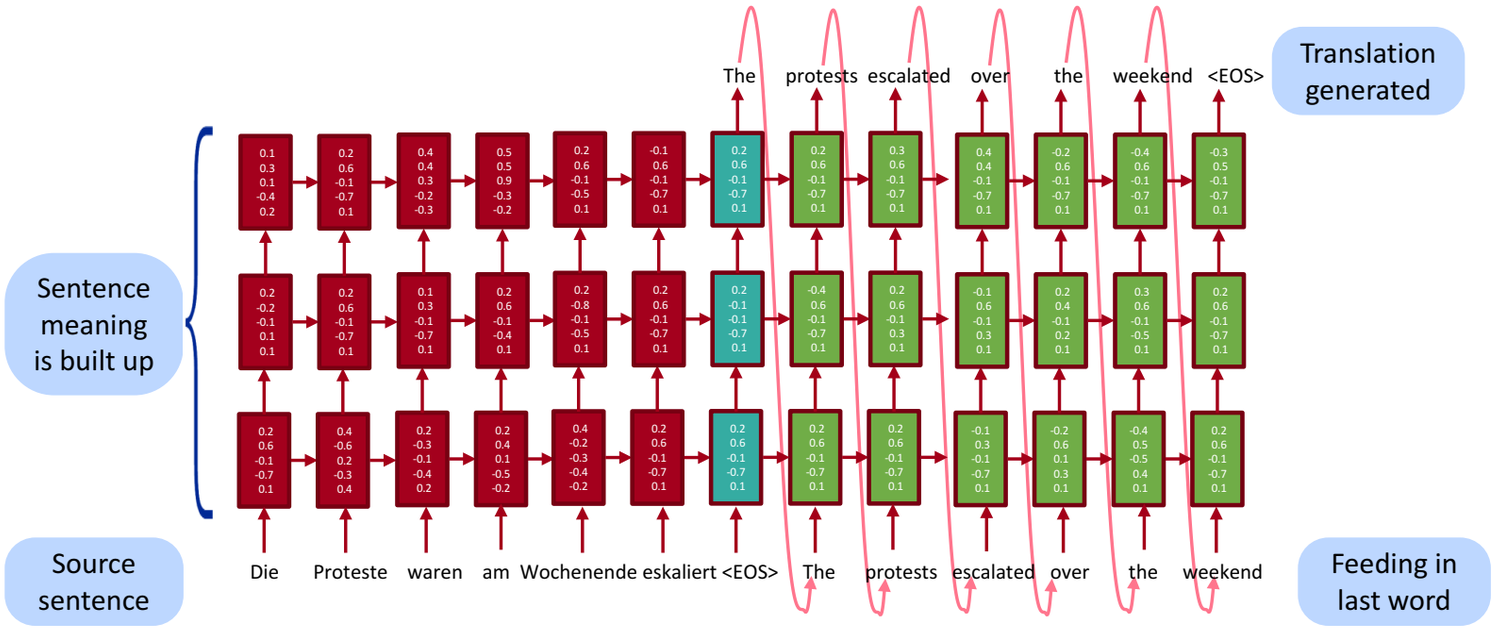
Zeiler and Fergus (2013)

Source: Stanford 224n

Deep Learning in Language

NTU NLP

Machine Translation: first breakthrough in 2015



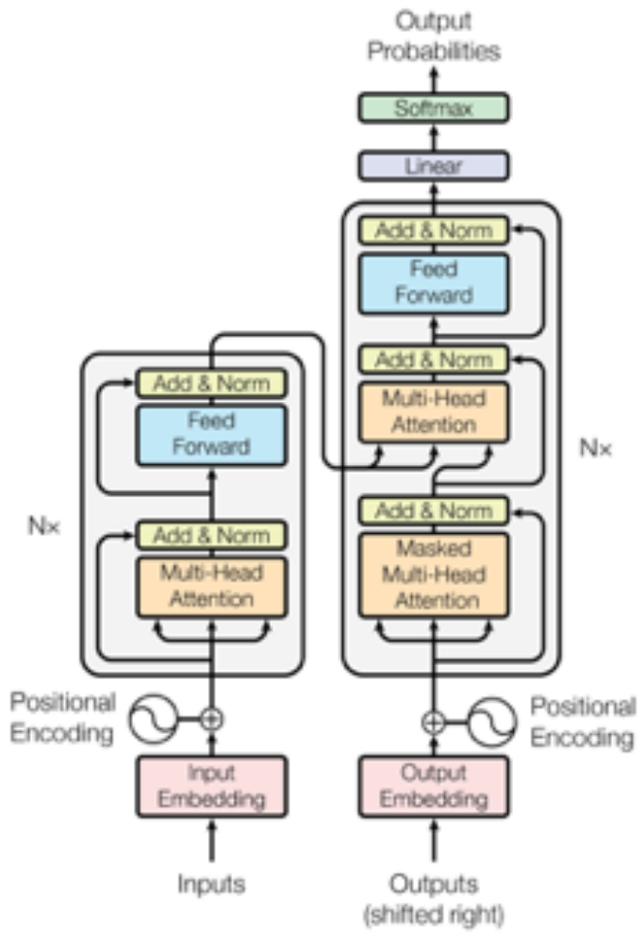


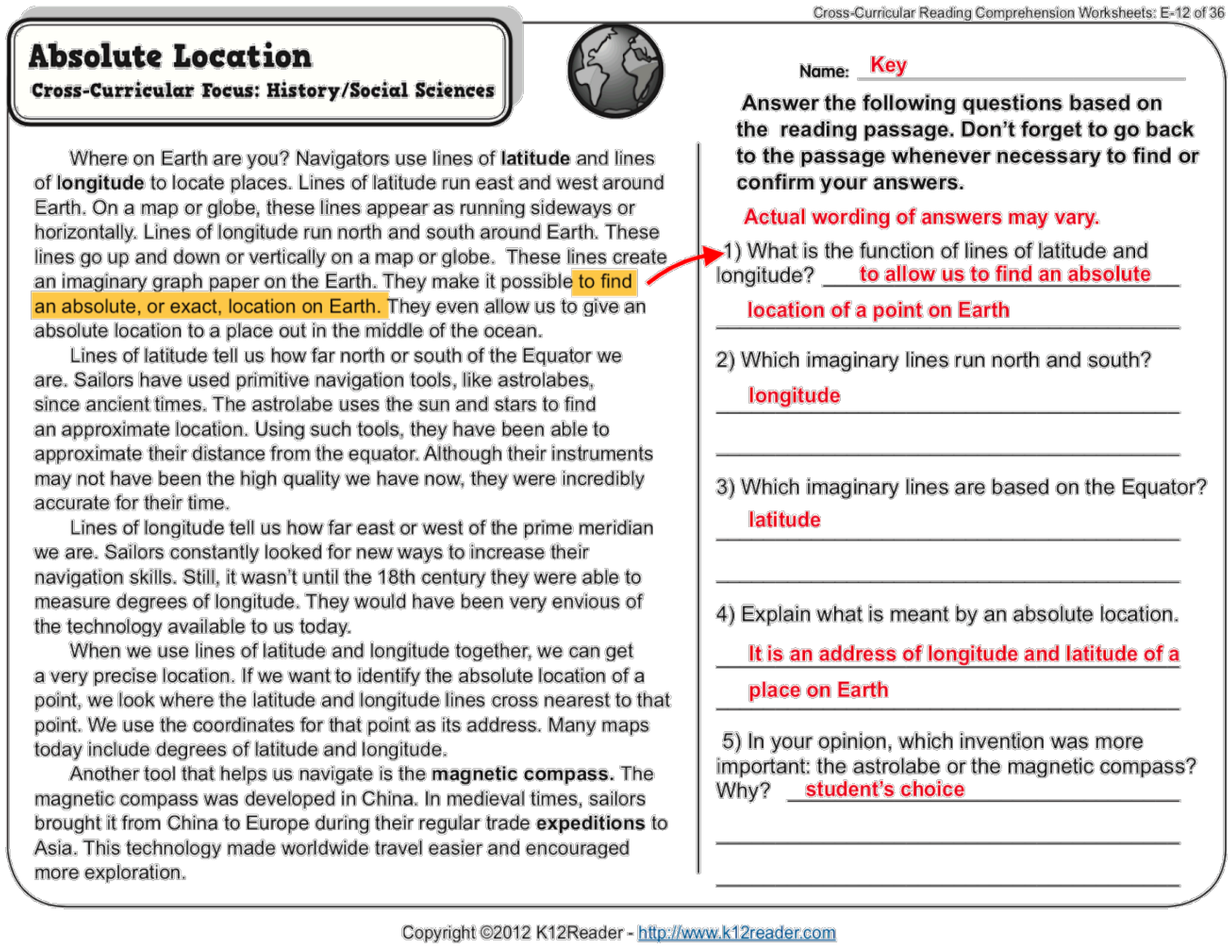
Figure 1: The Transformer - model architecture.

Now live for many languages in Google Translate (etc.), with big error reductions!

Deep Learning in Language

NTU NLP

Machine Reading Comprehension (QA)



Cross-Curricular Reading Comprehension Worksheets: E-12 of 36

Absolute Location

Cross-Curricular Focus: History/Social Sciences

Where on Earth are you? Navigators use lines of **latitude** and lines of **longitude** to locate places. Lines of latitude run east and west around Earth. On a map or globe, these lines appear as running sideways or horizontally. Lines of longitude run north and south around Earth. These lines go up and down or vertically on a map or globe. These lines create an imaginary graph paper on the Earth. They make it possible to find an absolute, or exact, location on Earth. They even allow us to give an absolute location to a place out in the middle of the ocean.

Lines of latitude tell us how far north or south of the Equator we are. Sailors have used primitive navigation tools, like astrolabes, since ancient times. The astrolabe uses the sun and stars to find an approximate location. Using such tools, they have been able to approximate their distance from the equator. Although their instruments may not have been the high quality we have now, they were incredibly accurate for their time.

Lines of longitude tell us how far east or west of the prime meridian we are. Sailors constantly looked for new ways to increase their navigation skills. Still, it wasn't until the 18th century they were able to measure degrees of longitude. They would have been very envious of the technology available to us today.

When we use lines of latitude and longitude together, we can get a very precise location. If we want to identify the absolute location of a point, we look where the latitude and longitude lines cross nearest to that point. We use the coordinates for that point as its address. Many maps today include degrees of latitude and longitude.

Another tool that helps us navigate is the **magnetic compass.** The magnetic compass was developed in China. In medieval times, sailors brought it from China to Europe during their regular trade **expeditions** to Asia. This technology made worldwide travel easier and encouraged more exploration.

Name: Key

Answer the following questions based on the reading passage. Don't forget to go back to the passage whenever necessary to find or confirm your answers.

Actual wording of answers may vary.

1) What is the function of lines of latitude and longitude? to allow us to find an absolute location of a point on Earth

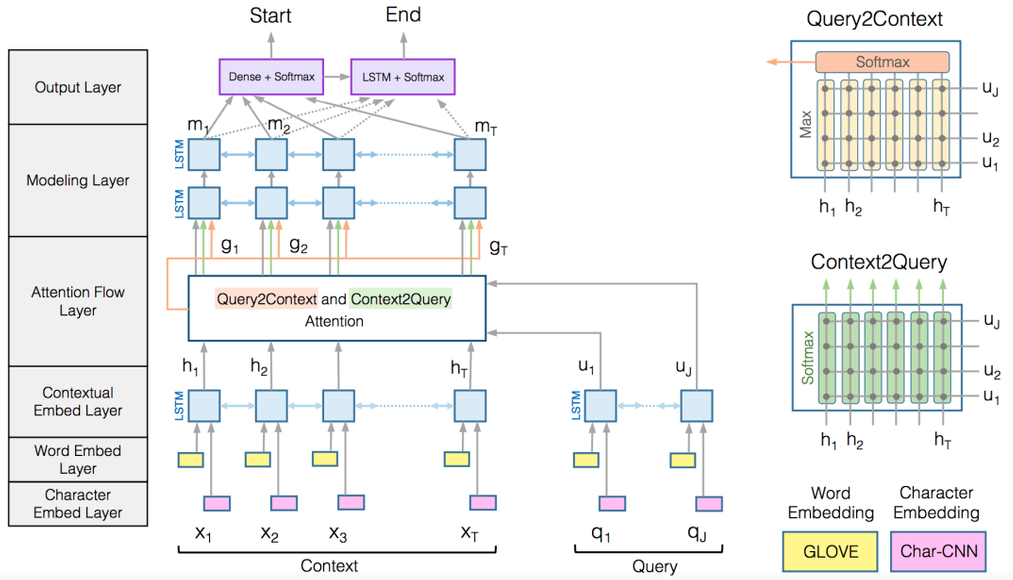
2) Which imaginary lines run north and south? longitude

3) Which imaginary lines are based on the Equator? latitude

4) Explain what is meant by an absolute location. It is an address of longitude and latitude of a place on Earth

5) In your opinion, which invention was more important: the astrolabe or the magnetic compass? Why? student's choice

Copyright ©2012 K12Reader - http://www.k12reader.com



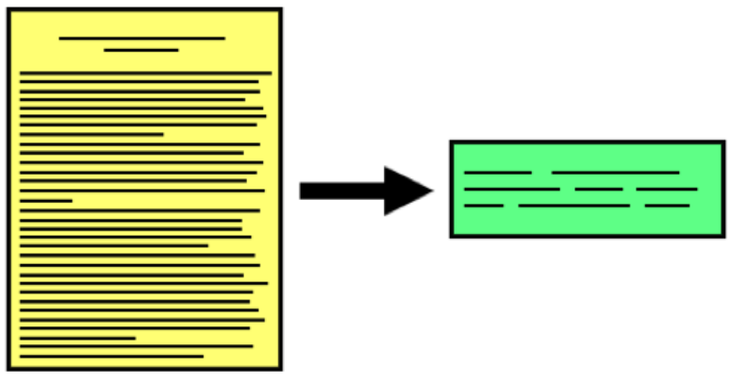
Allen AI's BiDAF

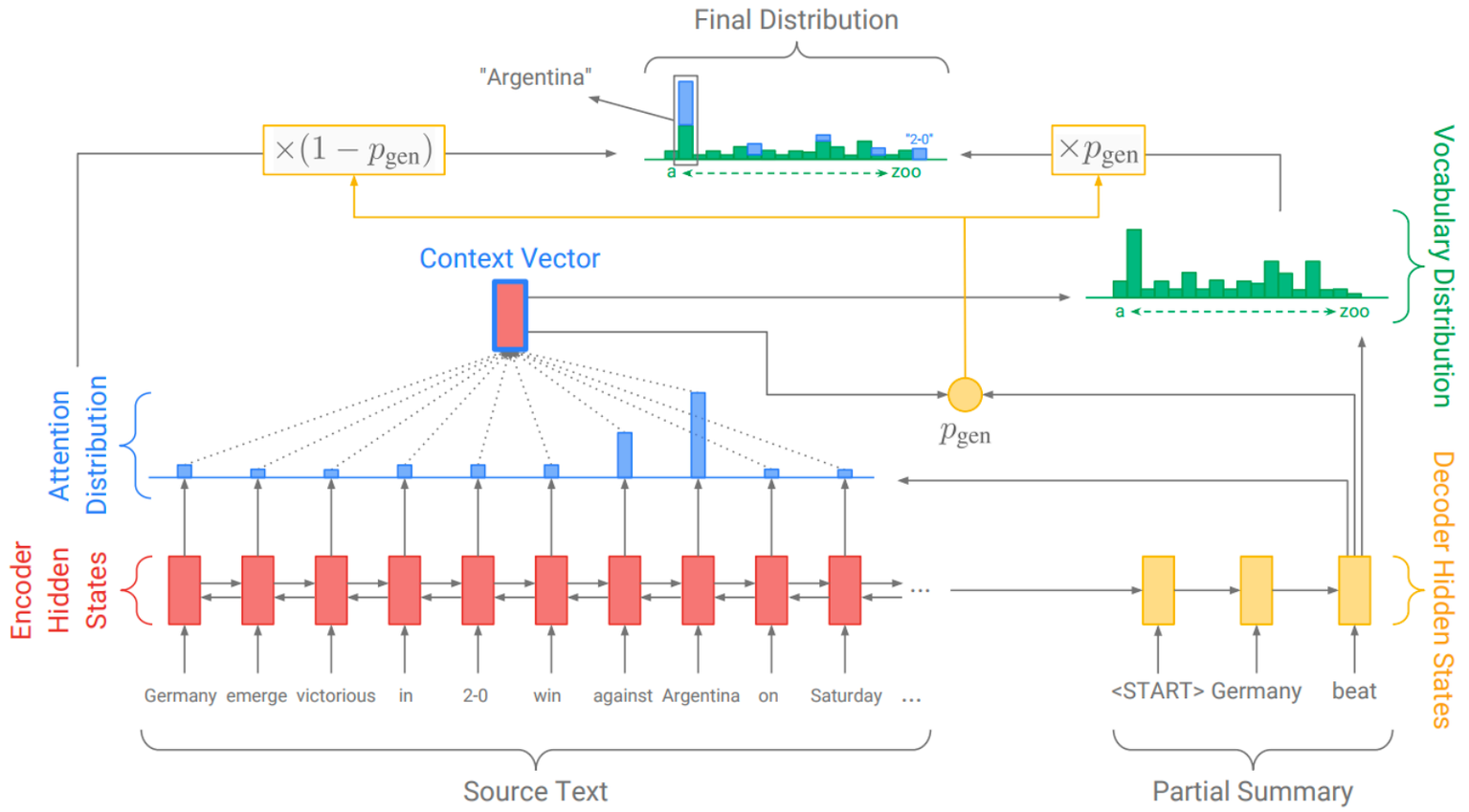
Source: Allen AI

Deep Learning in Language

NTU NLP

Text Summarization





Pointer Generator

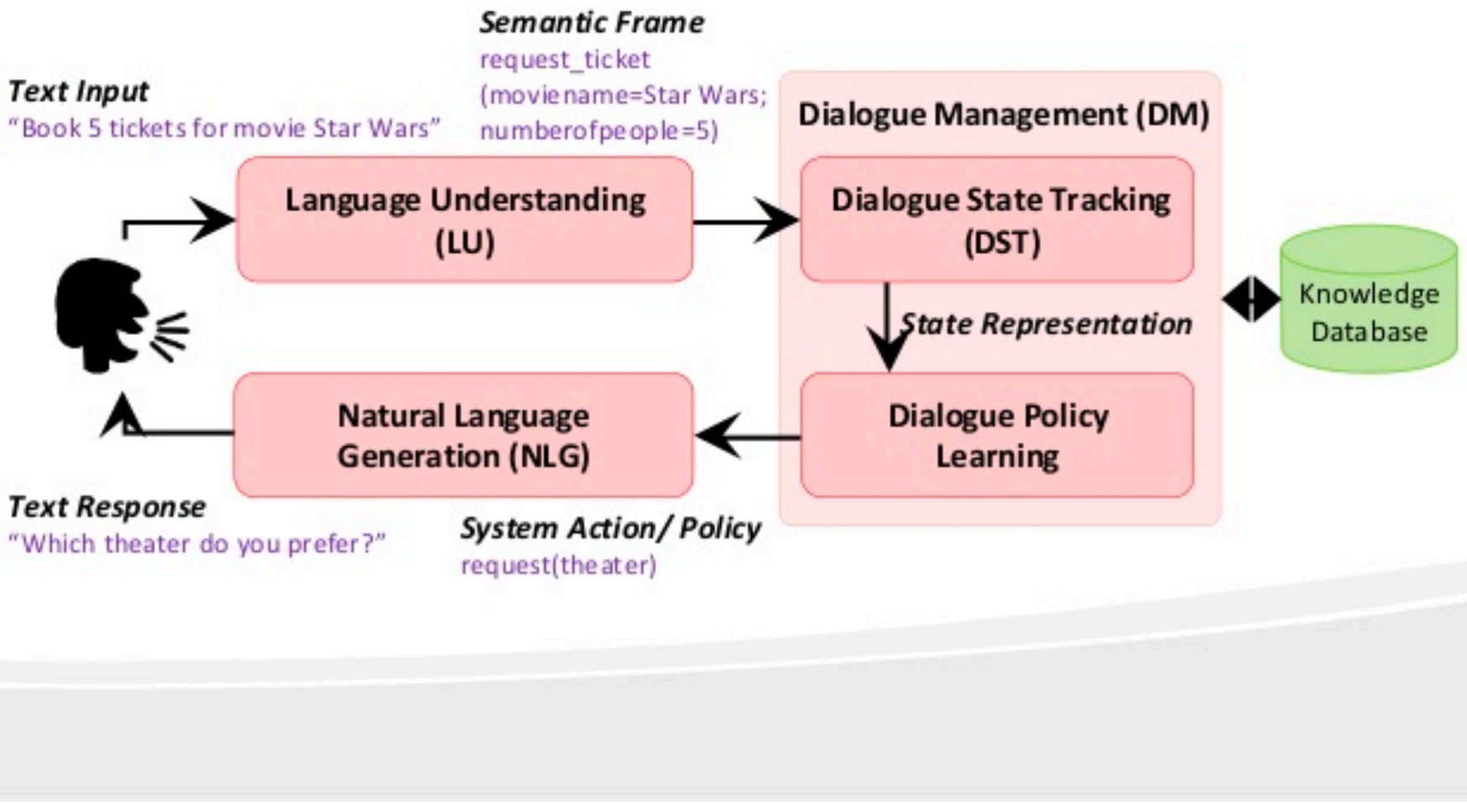
Source: See et al (2017)

Deep Learning in Language

NTU NLP

Dialogue Systems

9 Task-Oriented Dialogue System Framework



Source: MSRA

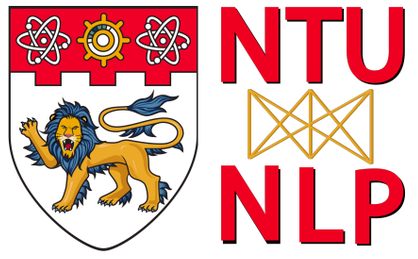
Why is Language Hard?

NTU NLP

- Complexity in representing, learning and using linguistic/situational/contextual/world/visual knowledge
- But interpretation depends on these
- Human languages are ambiguous (unlike programming and other formal languages)
- Examples:
  - I made her duck
  - Girl hit by car in hospital
  - Singapore Court to Try Shooting Defendant
  - Farmer Bill Dies in House

Source: Stanford 224n

DEEP NLP



Combine NLP knowledge with representation learning and deep learning

Improvements in recent years in all levels

- **Linguistic levels**: speech, words, syntax, semantics, discourse

- **Full applications**: sentiment analysis, question answering, dialogue agents, machine translation

Source: Stanford 224n

Deep Learning in Language

NTU NLP

Word Meaning

expect =

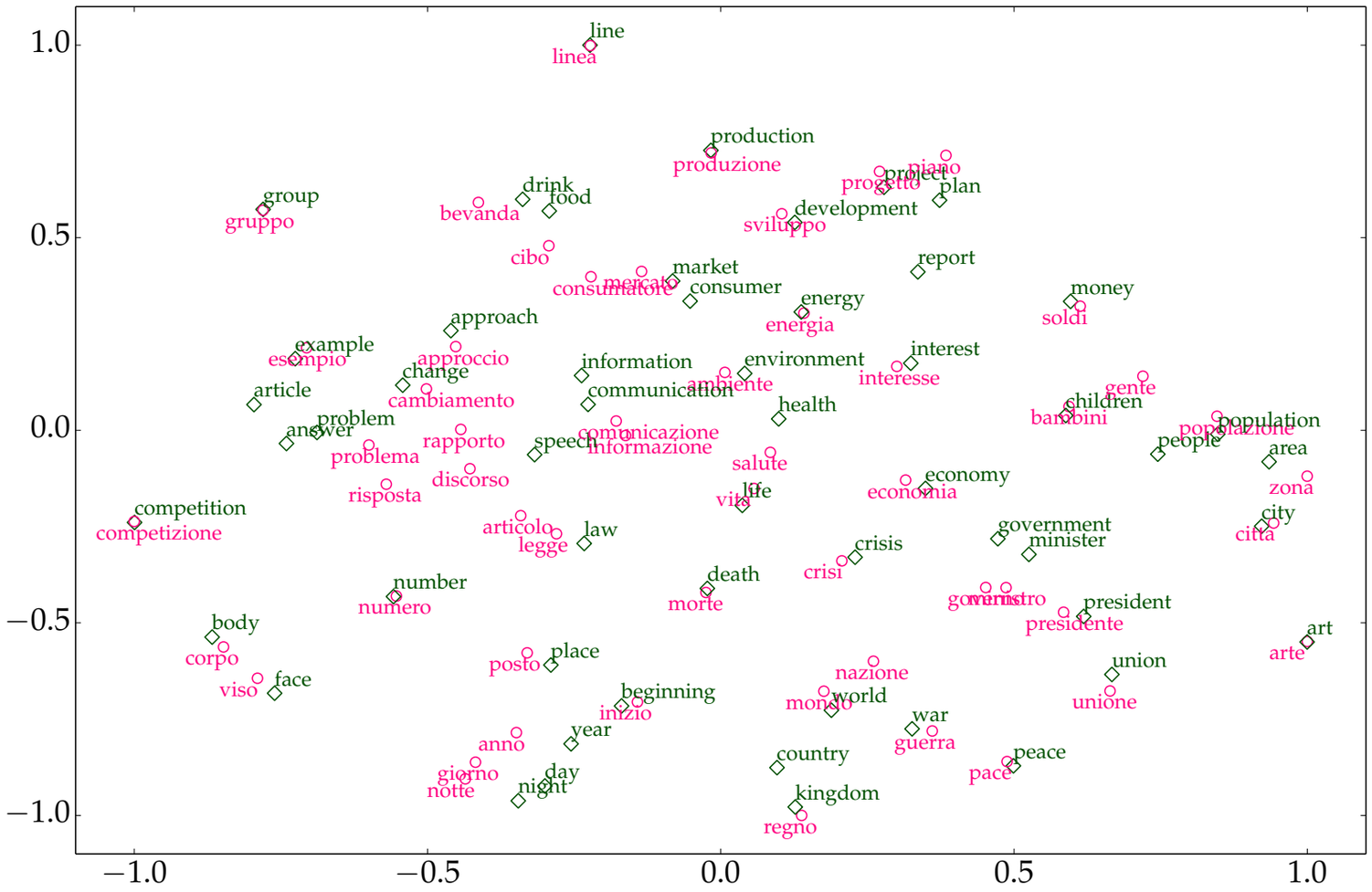
$$\begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \\ 0.487 \end{pmatrix}$$

Source: Stanford 224n

Deep Learning in Language

NTU NLP

Word Meaning (Cross-lingual)



Cross-lingual word embeddings (Ruder et al, 2019)

Source: Stanford 224n

Deep Learning in Language

NTU NLP

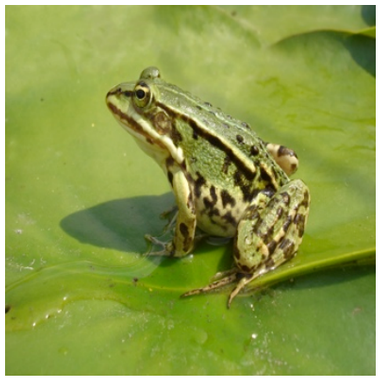
Word Similarities

Nearest words to frog:

1. frogs
2. toad
3. litoria
4. leptodactylidae
5. rana
6. lizard
7. eleutherodactylus

litoria

leptodactylidae



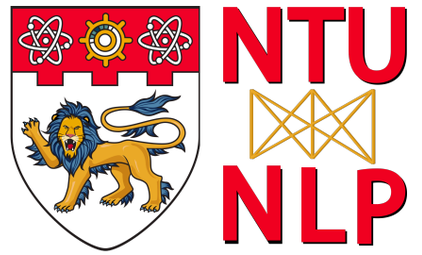
rana

eleutherodactylus

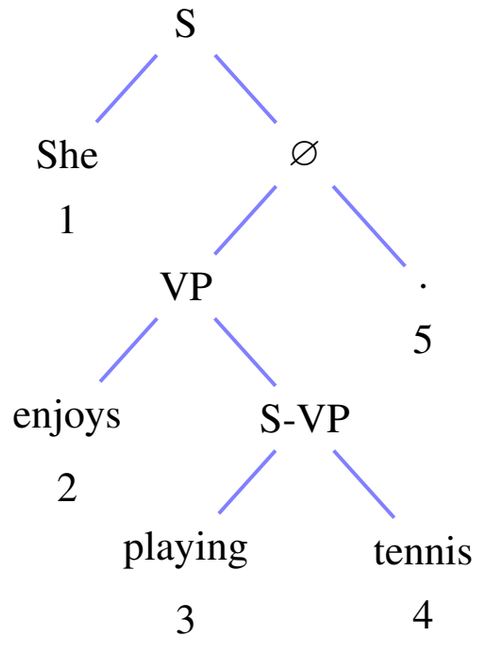
http://nlp.stanford.edu/projects/glove/

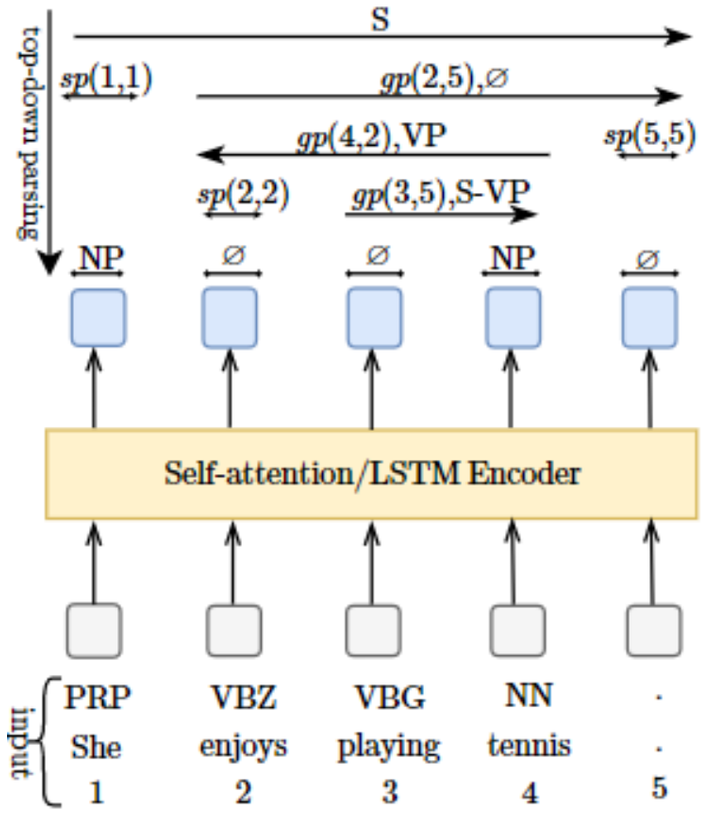
Source: Stanford 224n

Deep Learning in Language



Syntax (Parsing)





Neural networks can accurately determine the grammatical structure of sentences

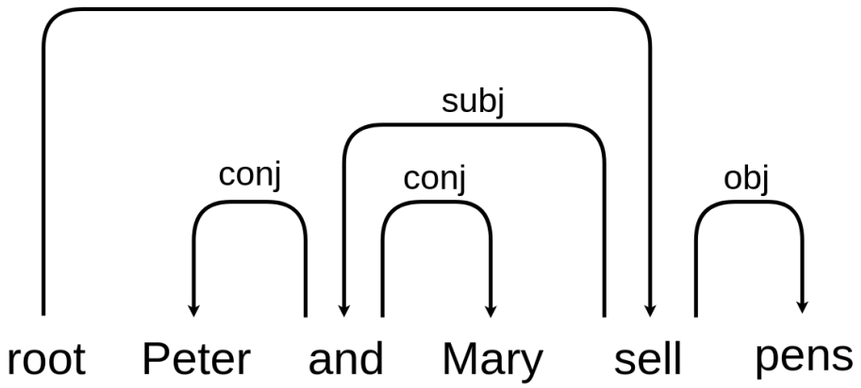
This supports interpretation and may help in disambiguation

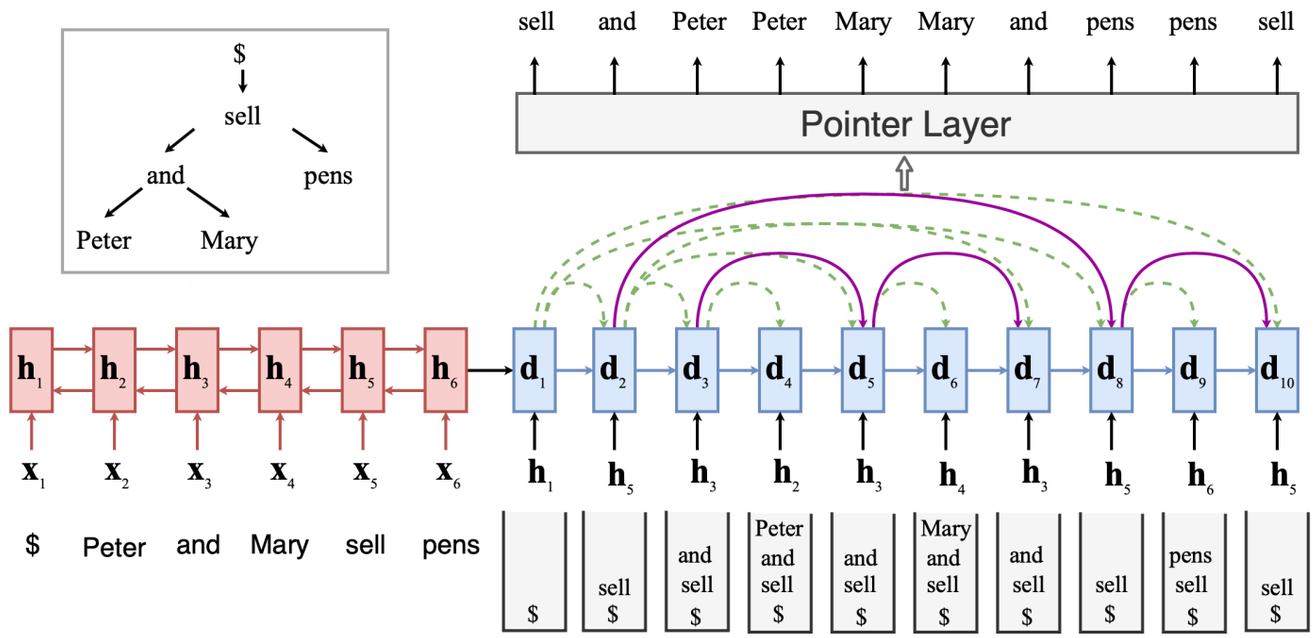
Source: NTU-NLP

Deep Learning in Language

NTU NLP

Syntax (Parsing)





Neural networks can accurately determine the grammatical structure of sentences

This supports interpretation and may help in disambiguation

Source: NTU-NLP

Deep Learning in Language

NTU NLP

Semantic Tagging

Named Entity Recognition (NER): Identify and classify entities such as persons, locations and organisations.

In 1917, Einstein applied the general theory of relativity to model the large-scale structure of the universe. He was visiting the United States when Adolf Hitler came to power in 1933 and did not go back to Germany, where he had been a professor at the Berlin Academy of Sciences. He settled in the U.S., becoming an American citizen in 1940. On the eve of World War II, he endorsed a letter to President Franklin D. Roosevelt alerting him to the potential development of "extremely powerful bombs of a new type" and recommending that the U.S. begin similar research. This eventually led to what would become the Manhattan Project. Einstein supported defending the Allied forces, but largely denounced using the new discovery of nuclear fission as a weapon. Later, with the British philosopher Bertrand Russell, Einstein signed the Russell-Einstein Manifesto, which highlighted the danger of nuclear weapons. Einstein was affiliated with the Institute for Advanced Study in Princeton, New Jersey, until his death in 1955.

Tag colours:
LOCATION TIME PERSON ORGANIZATION MONEY PERCENT DATE

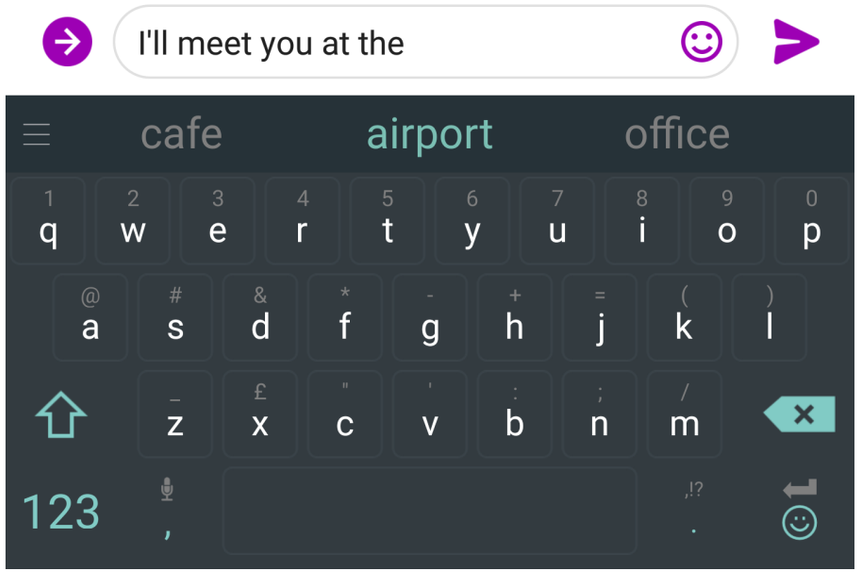
Deep neural networks achieve state-of-the-art results

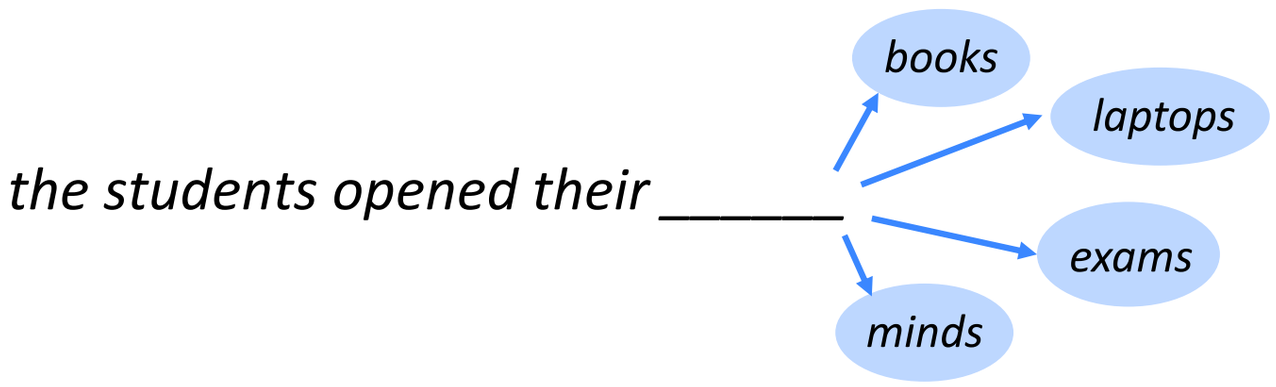
Deep Learning in Language

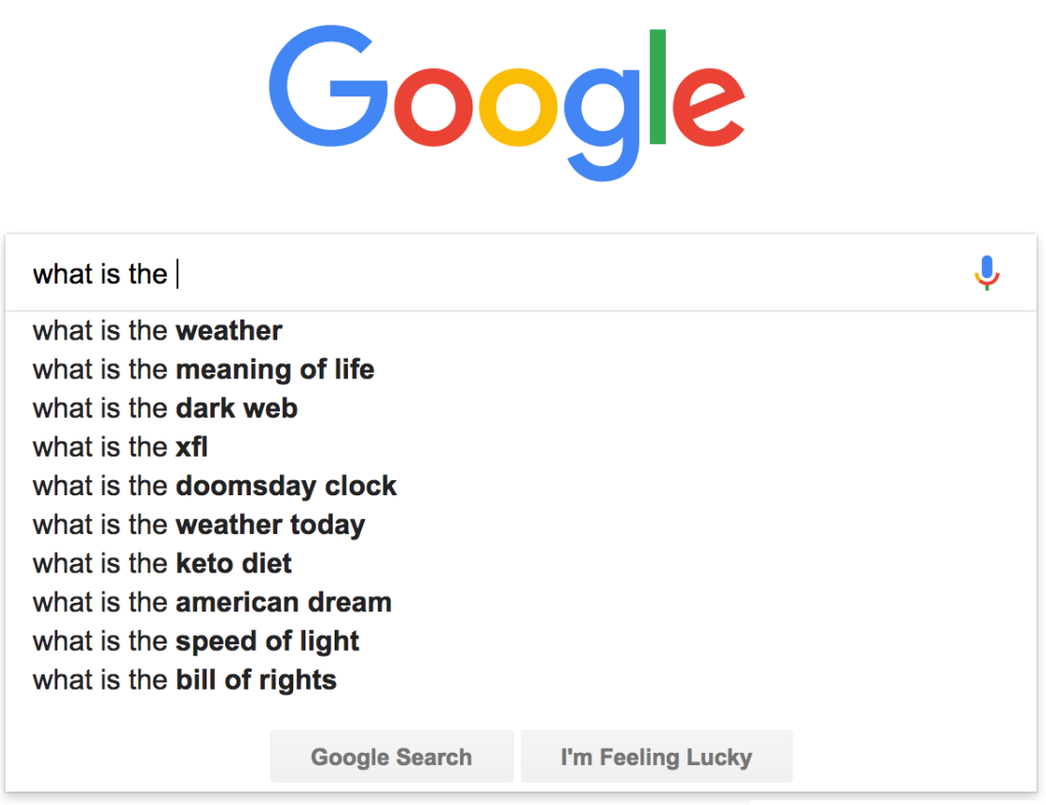
NTU NLP

Language Modelling

A language model takes a list of words (history/context), and attempts to predict the word that follows them







Deep neural networks have achieved state-of-the-art results

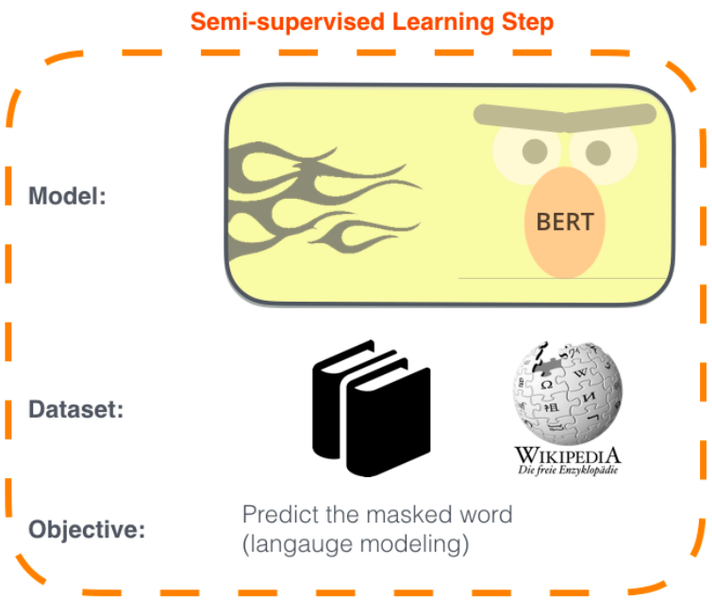
Deep Learning in Language

NTU NLP

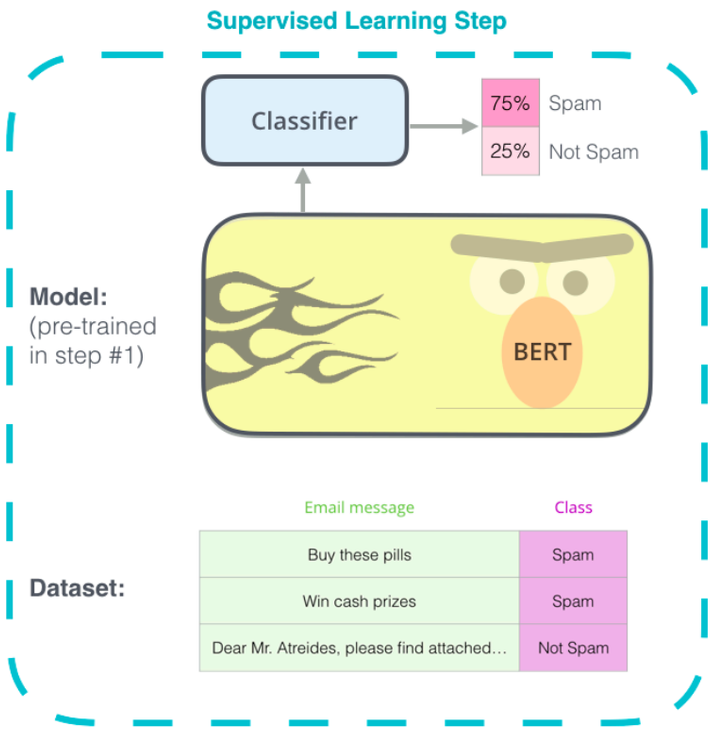
Representation Learning

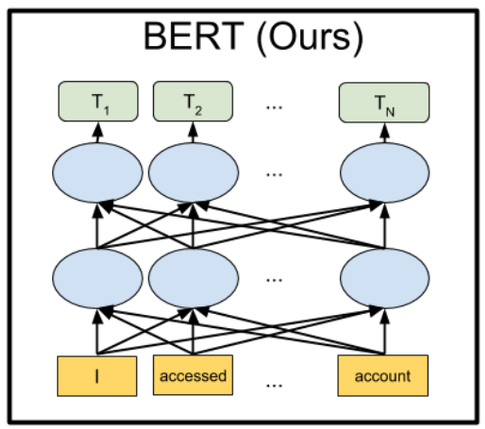
1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

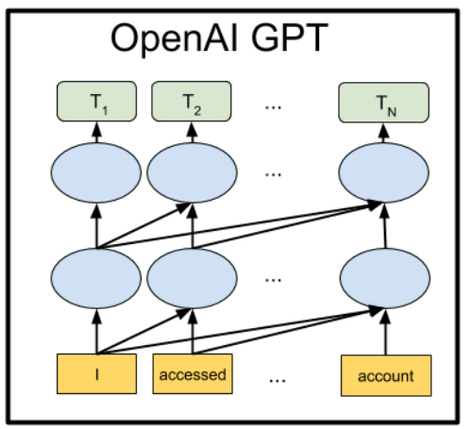
The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

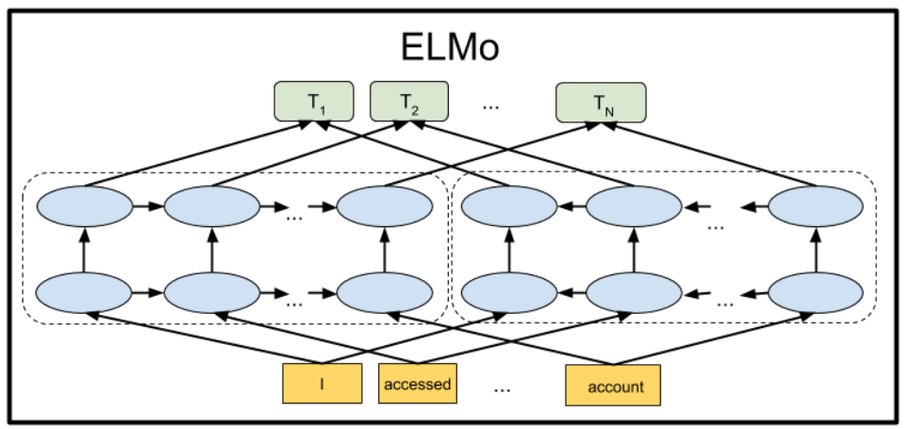


2 - Supervised training on a specific task with a labeled dataset.



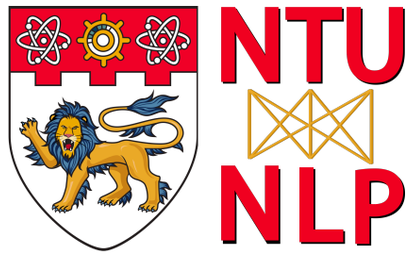






Source: http://jalammar.github.io/illustrated-bert/

Lecture Plan



- Course logistics
- What is NLP?
- What is Deep Learning
- Why deep learning for NLP
- Outline
- Knowing the target group

Outline

Lecture 1

- Introduction
- ML Basics

Lecture 2

- Feed-forward NN
- Word vectors

Lecture 3

- Recurrent NN
- Language models

Lecture 4

- Seq2Seq models
- MT, Summarization

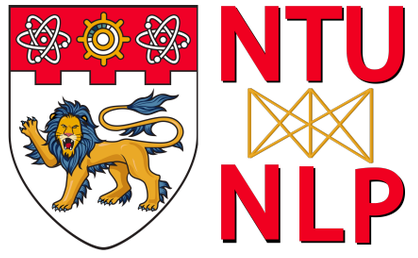
Lecture 5

- Seq2Seq with Attentions
- Transformers

Lecture 6

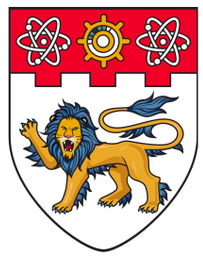
- Self-supervised learning
- Representation learning (BERT)

Slido QA



Slido channel: #94155

Knowing the Target Group



NTU NLP

- Do you know Python?
- Do you know Pytorch?
- Do you know Machine Learning?
- Do you know Deep Learning?
- Have you taken a course in AI?
- Have you taken a course in NLP?
- What is your expectation from this course?

Slido channel: #94155