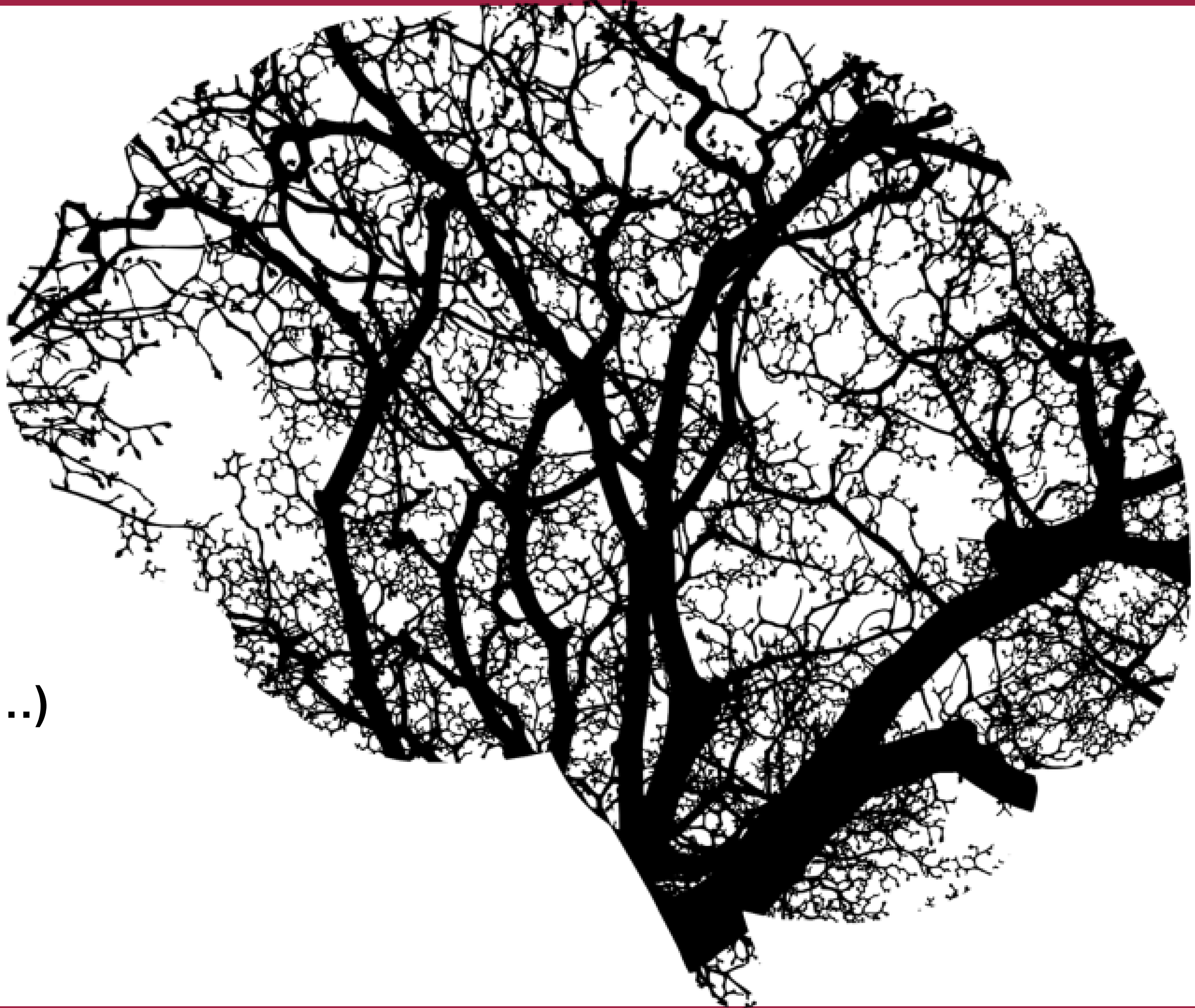
enliteAI
one neuron at a time



NLP - State of the Art
(and some missing parts…)

Liad Magen

Wien, am 27.03.2019

Hello, World!

Israel

computer science

biomedical engineering

NLP & AI Researcher @ enliteAI

Keep-Current Project:

- Machine Learning Seminar
- fast.ai + PyTorch course

data4good

Services by enliteAI

AI Strategy & Transformation

- Conduct AI readiness check across organization
- Discover Use-cases to build-up momentum
- Define AI Strategy

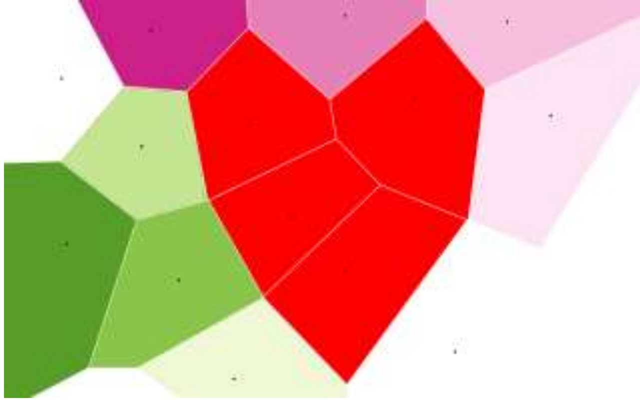
AI Lab

- Leverage AI Research for practical applications
- Extend research to production-grade
- Build models for commercial use

Prototyping & Project Delivery

- Development of POCs, MVPs and prototypes
- Implementation & Integration
- Roll-out & Support

data4good Hackathon



Applying data science to human needs and generating social impact:

- GrünStattGrau
- Hilfswerk International
- Hilfswerk Österreich
- CivesSolutions

April 27-28th, 2019

A1 Telekom, Lassallestraße 9

What is NLP?

Enabling computers to interact with users in their language

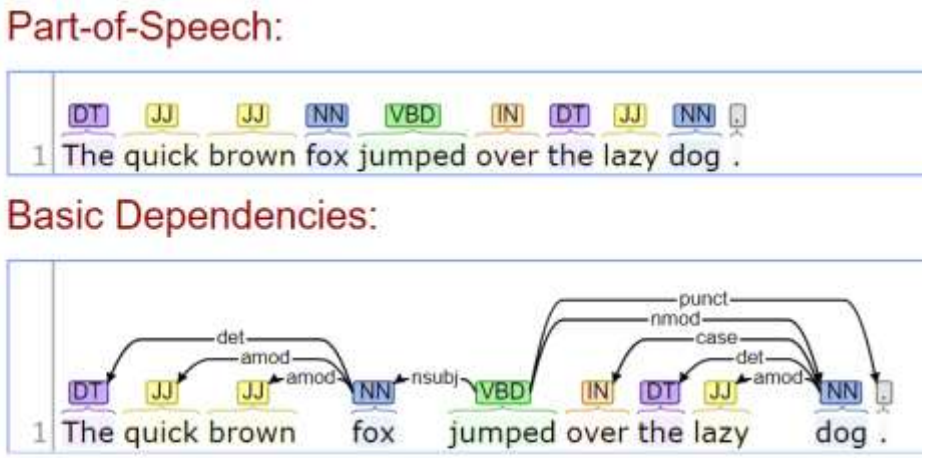
NLP is everywhere

Applications:

- Language understanding
- Search engines
- Machine translation
- Text analysis (i.e. sentiment)
- Chatbots / Personal Assistance

NLP Building blocks

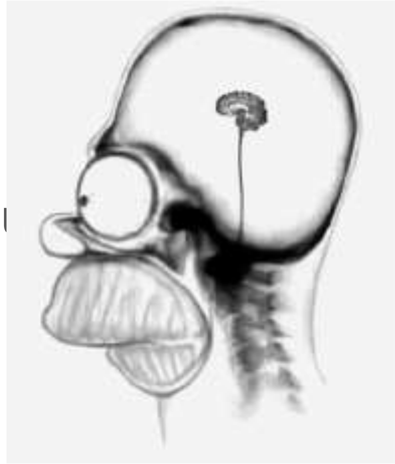
- Chunking
- Sequence Tagging
  - Part of speech
  - Named Entity Recognition (NER)
  - Coreference Resolution
  - ...
- Syntactic Parsing
- And more...



NLP in a nutshell

Human Language

Non-trivial useful ou



takes as input text in human language and process it
in a way that suggests an intelligent process was involved

NLP is hard!

I’m eating pizza with olives

I’m eating pizza with friends

NLP is hard!

Google translate:

State of the art

German

Der letzte Stand der Technik

Hebrew

State of the art

“Country of art”

Variability

Ambiguity

He acquired it
He purchased it
He bought it
It was bought by him
It was sold to him
She sold it to him
She sold him that

Twitter:
I loooooove it!
everytime → every time
oscar nom’d doc → Oscar-nominated documentary
is bananas → is great

Bank

Apple

Star

Spring

Play

I’m reading a **Book**

I will **Book** the flight

- by **Wei Xu**

Ambiguity - Real news headlines

- Scientists examine whales from space
- The pope's baby step on gays
- Girl hit by car in hospital
- Iraqi head seeks arms
- Lung cancer in women mushrooms
- Squad helps dog bite victim
- Miners refuse to work after death
- Stolen painting found by tree
- Actor sent to jail for not finishing sentence
- Stadium air conditioning fails — Fans protest

What do you mean no?

Martin is a data scientist

Martin is a great data scientist

What do you mean no?

Martin is a data scientist

Martin is **not** a data scientist

Martin is a great data scientist

Martin is **not** a great data scientist

What do you mean no?

Martin is a data scientist
Martin is **not** a data scientist

Martin is a great data scientist
Martin is **not** a great data scientist

Martin is **not** a data scientist
Martin is **not** a mad scientist

Negation + Monotonicity

John did not work for Microsoft

in 1989

as a VP

of sales

Restrictivity / Intersectivity

He is a great pianist → He is a pianist

He is a fake doctor → He is a doctor

I found my old jacket → I found my jacket

How do we do NLP?

Linguistics Experience

Rule based Systems
1950s-1990s

Corpus based statistics
1990s-2000s

Machine Learning
2000s-2014

Deep Learning
2014-2020(?)

Machine Learning Expertise

Transparent

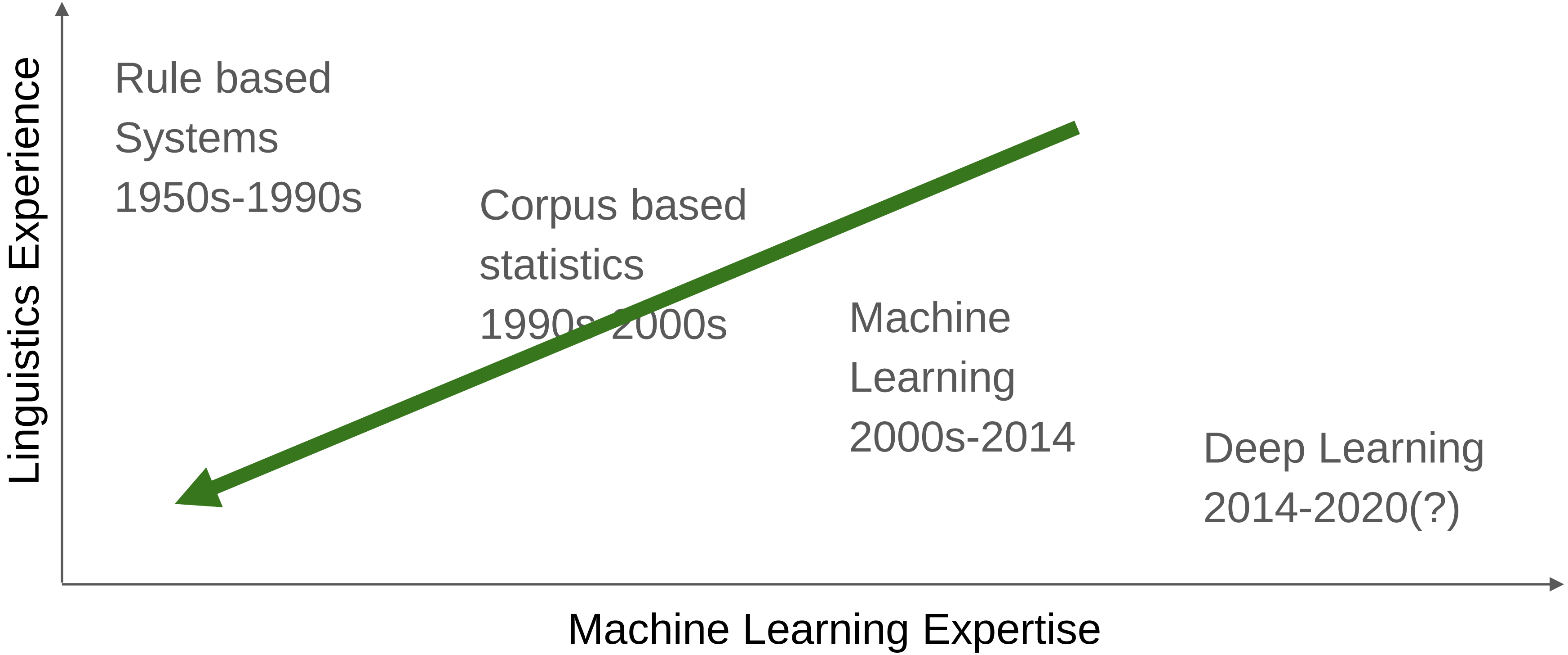
Debugability

Black box

Little

Labeled Data

How should we do NLP?



Transparent

Debugability

Black box

Little

Labeled Data

NLP Tomorrow

Linguistics Experience

Rule based
Systems
1950s-1990s

Corpus based
statistics
1990s-2000s

Machine
Learning
2000s-2014

Deep Learning
2014-2020(?)

Humans writing rules aided by ML/DL - Resulting in transparent and debuggable models

Machine Learning Expertise

Transparent

Debugability

Black box

Little

Labeled Data

NLP Today

3. The BiLSTM Hegemony

To a first approximation,
the de facto consensus in NLP in 2017 is
that no matter what the task,
you throw a BiLSTM at it, with
attention if you need information flow

28

- Chris Manning (Stanford) @ Simon’s institute - 27.3.2017

BiLSTM hagmoney (> 2015)

Works on low level tasks:

- Coordination boundary prediction:
  He will attend the meeting **and** present the results on Tuesday
  He will attend the meeting on Tuesday
  He will present the results on Tuesday
- Syntactic Parsing
- And more...

BiLSTM hagmoney (> 2015)

But less on higher levels:

BiLSTM text generation with style:

Parameter	Value
Theme	Other
Sentiment	Positive
Professional	True
Personal	False
Length	11-20 words
Descriptive	False

“The film’s ultimate pleasure if you want to fall in love with the ending, you won’t be disappointed”
“The film’s simple, and a refreshing take on the complex family drama of the regions of human intelligence.”

https://arxiv.org/pdf/1707.02633.pdf - Controlling Linguistic Style Aspects in Neural Language Generation

BiLSTM hagmoney (> 2015)

But less on higher levels:

BiLSTM text generation with style:

Parameter	Value
Theme	Other
Sentiment	Negative
Professional	False
Personal	True
Length	11-20 words
Descriptive	True

“There are some funny parts but overall I didn’t like the first few funny parts, but overall pretty decent .”

“Ultimately, I can honestly say that this movie is full of stupid stupid and stupid stupid stupid stupid stupid.”

https://arxiv.org/pdf/1707.02633.pdf - Controlling Linguistic Style Aspects in Neural Language Generation

NLP in the academy

Multiple stacked networks

Each architecture solves a ***specific*** problem
(or a dataset...)

NLP in the industry

NGrams, TF/IDF, LDA / Topic modeling

Word2vec / GloVe / FastText
ELMo (AI2) / BERT / ERNIE (Baidu)
GPT-2 (OpenAI)

LSTM

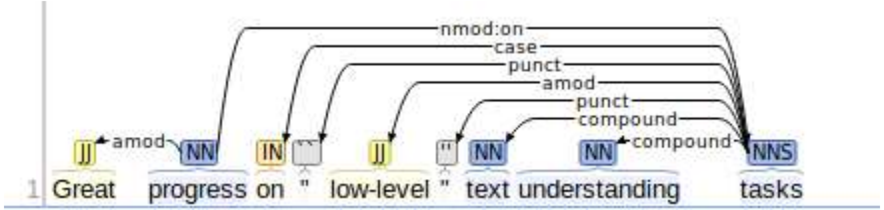
spaCy
Flair

& Regular Expressions...

20 years of NLP research

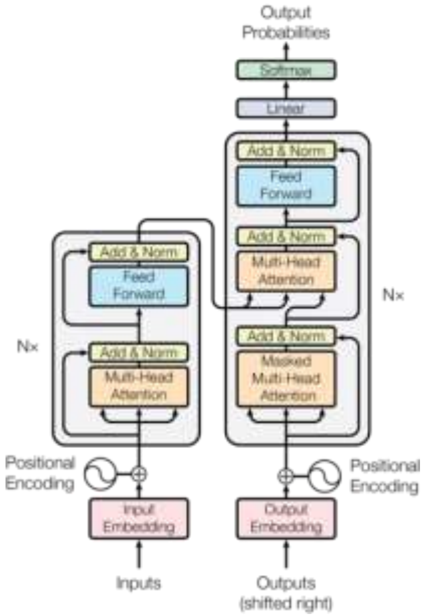
Great progress on “low-level” text understanding tasks:

- Morphology
- Syntax / Parsing
- Semantic Roles.
- Lexical Semantics
- Discourse. Coreference.
- Negation scope



Representation learning, text-to-text semantics, RNNs

- Word vectors
- Concept vectors
- Sentence vectors
- Paraphrasing



20 years of NLP research

Great progress on “low-level” text understanding tasks:

- Morphology
- Syntax / Parsing
- Semantic Roles.
- Lexical Semantics
- Discourse. Coreference.
- Negation scope

Remarkable Achievements

Not useful for non-experts

Representation learning, text-to-text semantics, RNNs

- Word vectors
- Concept vectors
- Sentence vectors
- Paraphrasing

Automatically learn intricate linguistic patterns

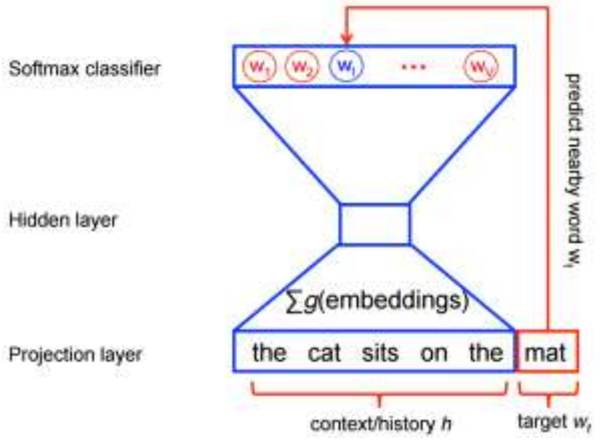
Bad with nuance.
Contains biases.
Hard to control.

Short recup of the last few years

2013

Language model

Learn joint likelihood of training sentences



Word2Vec

… GloVe … FastText …

First use of *transfer learning* in NLP

2018

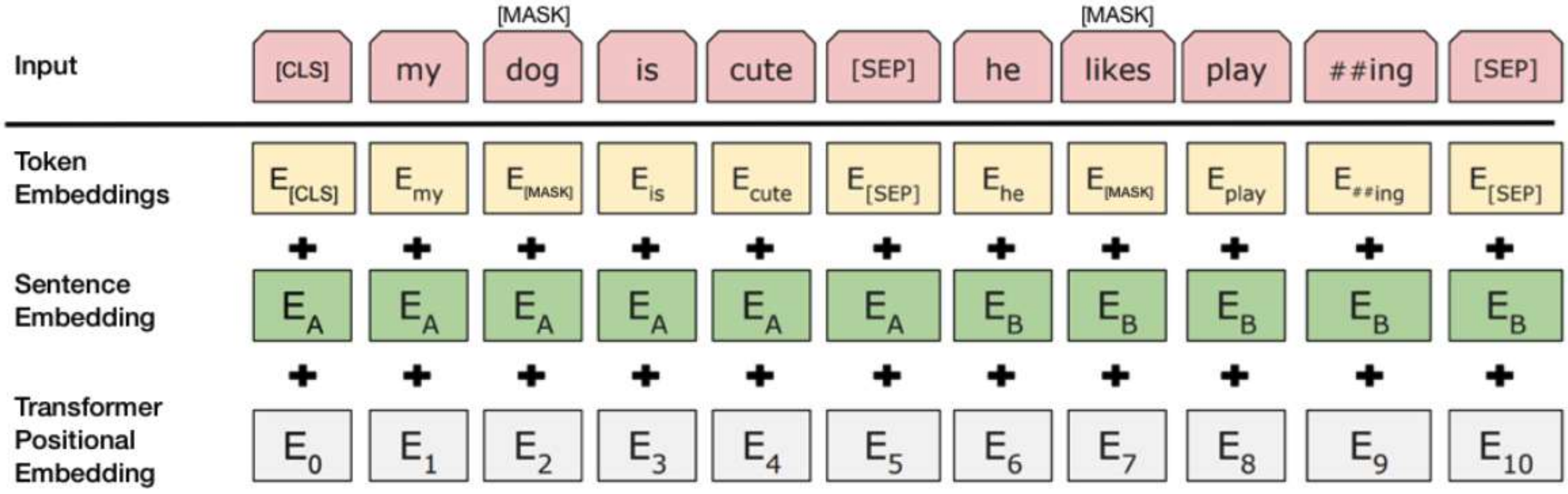
ELMo - Contextual word vectors
ULMFiT - transfer learning + fine tuning
BERT - the above & more, on large scale

2019

Transformer-XL
GPT-2

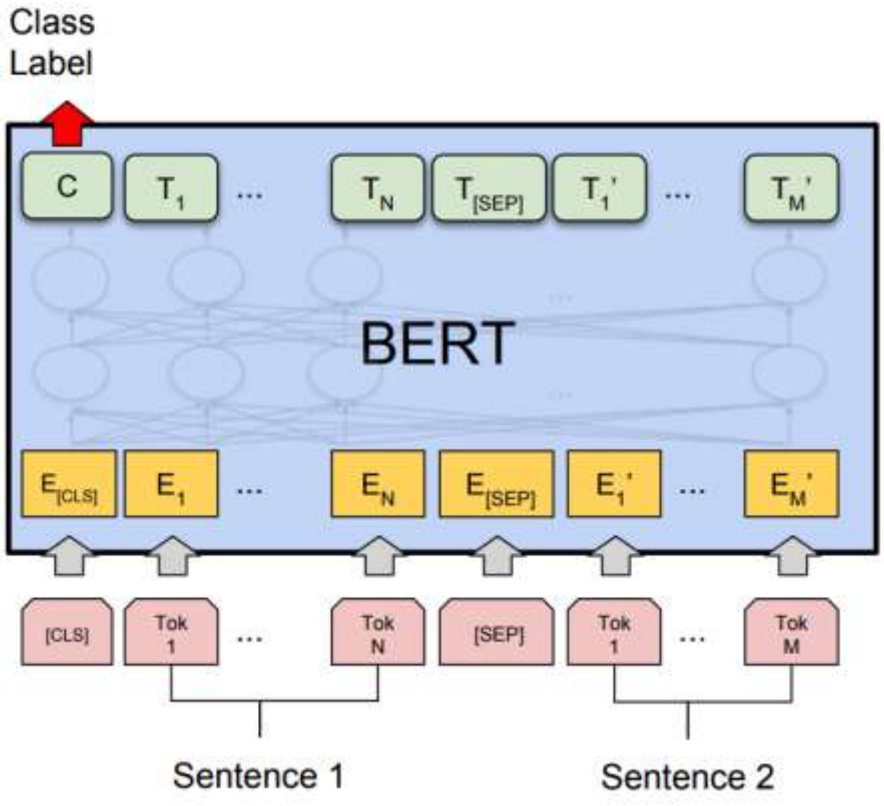
BERT - architecture in brief

Embedding:



BERT - architecture in brief

- Embedding + Masking (Cloze)
- Transformer
- Sentence pairing
- Fine tuning
  - Sentence class Label
  - NER Sequence Labels
- Question Answering:
  - Start+End Vector & Softmax over all positions (as additional parameters)
  - Or Running Q + A (for each option) and training on the class label

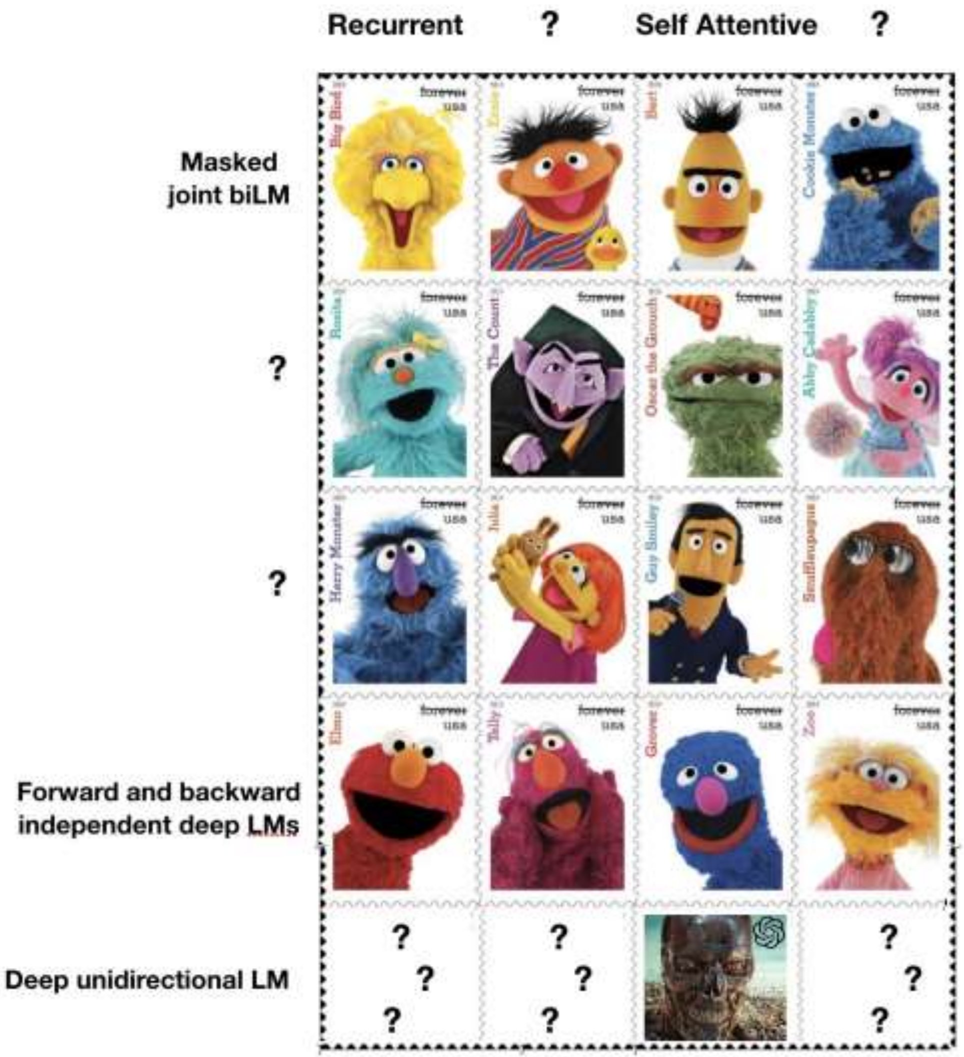


Further reading

- Toward data science:
  - Dissecting BERT (parts 1 & 2) - by Miguel Romero Calvo
  - Deconstructing BERT (parts 1 & 2) - by Jesse Vig
- HarvardNLP - The Annotated Transformer
- Jay Alammar - The illustrated BERT, ELMo and co.
- NLP.Stanford.edu/Seminar - Jacob Delvin presentation about BERT

Size matters

- ULMFiT
- ELMo
- GLoMo
- GPT
- BERT
- Transformer-XL
- MT-DNN
- GPT-2



- Contextual representations
- Generalized Language models
- Transfer learning in NLP

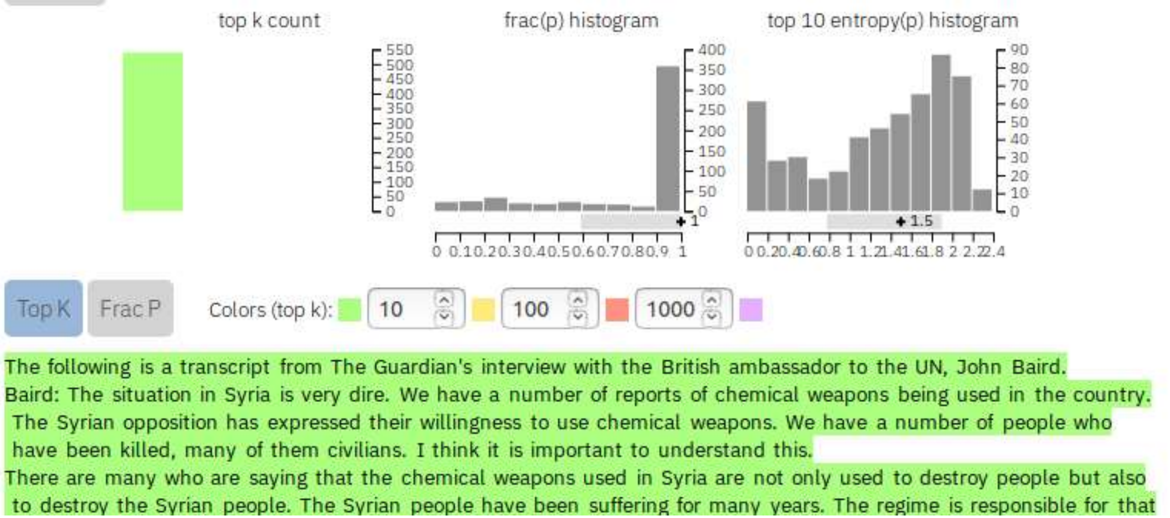
Text Generation

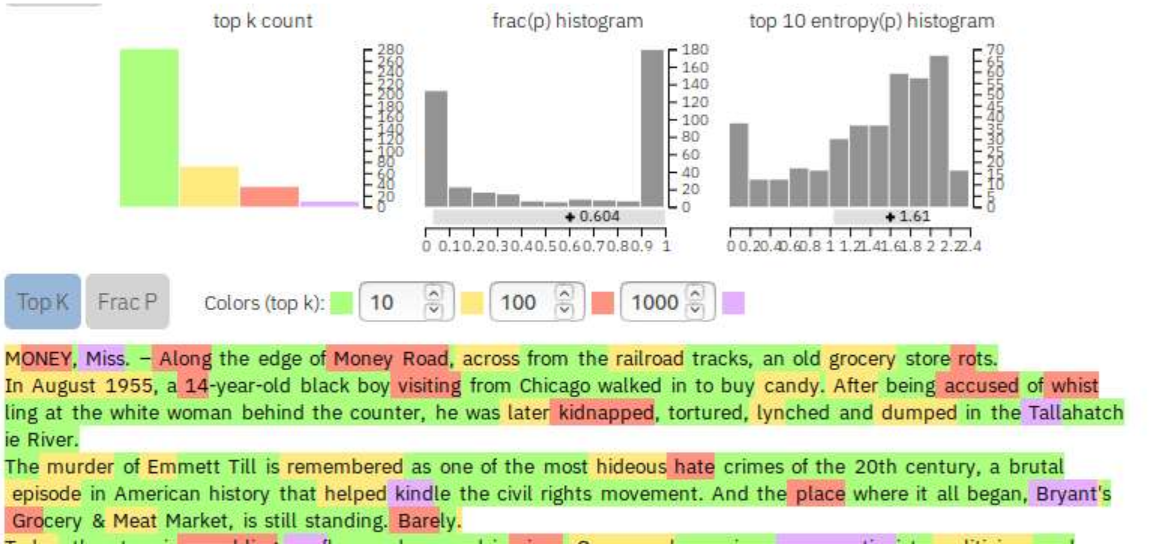
Dissecting GPT-2 results

HavardNLP

Giant **L**anguage model **T**est **R**oom

http://gltr.io





NLP Progress - sample domains

Cross lingual Text Classification: pyTorch - LASER

Text Classification (English): ULMFiT

Automatic Speech Recognition (ASR): Google Tensorflow Lingvo

Translation: DeepL / Transformer Big (Facebook + Google)

NER: flair / BERT

Entity linking: DeepType - OpenAI

nlpprogress.com | github.com/syhw/wer_are_we

Q&A Datasets & Benchmarks

General Language Understanding Evaluation (GLUE)

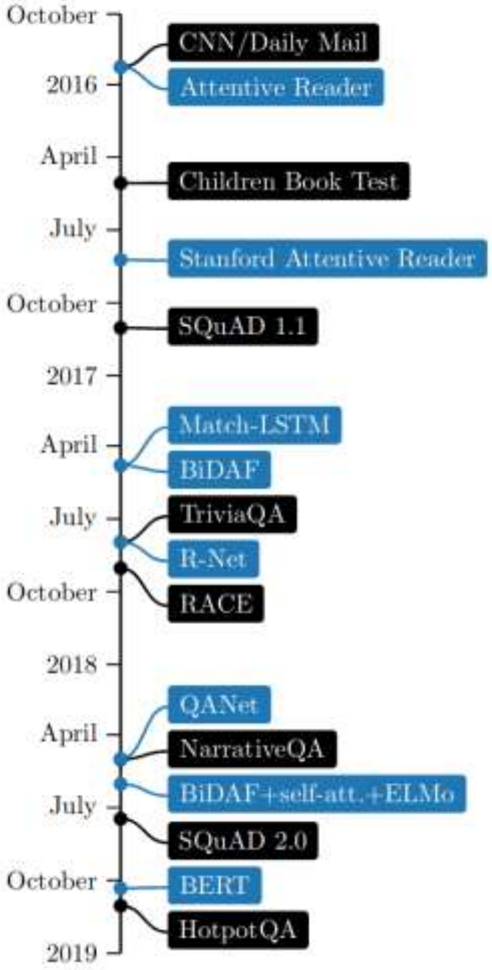
Stanford Natural Language Inference (SNLI)

SQuAD / SQuAD 2.0

RACE (Carnegie Mellon)

HotpotQA, MS Marco, NarrativeQA, CoQA ...

Google AI natural questions dataset



Question Answering - State of the art

Question Answering - State of the art?

Depends on the dataset…

Each information extraction (IE) task is a different NLP application

(Also relevant for chatbots / personal assistants)

Information Extraction

Turning text into structured data

- Most data in the world is in text
- We can only easily analyze structured data
- Today - no easy way to do it on scale
- Possible for **specific** tasks - but requires
    - Enough data
    - Team of NLP experts
    - Time

Solving IE can be transformative to industry & science

Question Answering - next steps

In which state was Yahoo founded?

History of Yahoo!

From Wikipedia, the free encyclopedia

See also: Timeline of Yahoo!

This article needs to be **updated**. Please update this article to reflect recent events or newly available information. *(May 2016)*

Yahoo! was started at Stanford University. It was founded in January 1994 by Jerry Yang and David Filo, who were Electrical Engineering graduate students when they created a website named "Jerry and David's Guide to the World Wide Web". The Guide was a directory of other websites, organized in a hierarchy, as opposed to a searchable index of pages. In April 1994, Jerry and David's Guide to the World Wide Web was renamed "Yahoo!".[1][2] The word "YAHOO" is a backronym for "Yet Another Hierarchically Organized Oracle"[3] or "Yet Another Hierarchical Officious Oracle."[4] The yahoo.com domain was created on January 18, 1995.[5]

http://ai.stanford.edu/blog/beyond-local-pattern-matching/

Question Answering - next steps

- (Cleverly) add more inductive biases
- Common sense

Possible methods:

- Interactive learning (human in the loop)
- Deep reinforcement learning
- GANs

Still missing...

1. Understanding *(Not only NLP-related…)*
    - What is captured by a network?
    - What can/’t the networks learn
2. Explainability
3. Fairness & accountability
    - Removing bias (current techniques only “put a lipstick on a pig”)
4. Working with little annotated data
   Transferring knowledge across *domains*
5. Handling missing data

Thank you!

We’re “standing on the shoulders of giants”

- Isaac Newton

enliteAI

Appendix 1 - handling missing data

Handling missing data

- Mary is a great programmer but John isn't
- The hotel received great reviews at Yelp but mediocre reviews at TripAdvisor
- Fred took a picture of you and Susan of me

OPEC agrees to cut oil production, defying Trump

Following two days of tense talks in Vienna, a deal to reduce oil production has been struck by OPEC and Russia. Prices have already risen on the back of the news, but Donald Trump will not be pleased.

Handling missing data

- Mary is a great programmer but John isn't _____ *(verb phrase ellipsis)*
- The hotel received great reviews at Yelp but _____ mediocre reviews at TripAdvisor *(argument clusters)*
- Fred took a picture of you and Susan ______ of me *(gapping)*

OPEC agrees to cut oil production, defying Trump

Following two days of tense talks in Vienna, a deal to reduce oil production has been struck by OPEC and Russia. Prices have already risen on the back of the news, but Donald Trump will not be pleased. **Prices of oil**

(Briding)

I turn 40 this year

He drove 40 on the highway

I’ll give you 40 for that

I turn 40 <u>(age)</u> this year

He drove 40 <u>(kmh)</u> on the highway

I’ll give you 40 <u>(currency)</u> for that

Handling missing data

- Current end-to-end systems fails on them
- Stanford parser can partially handle several cases
- Work in progress in BIU

Appendix 2
Transferring knowledge across domains

Example - review sentiment analysis

“The soup was delicious” → +

“The steak was tough” → -

“The dessert was decadent” → +

[food] **was** [good_food_adj]

[food] **was** [bad_food_adj]

Example - review sentiment analysis

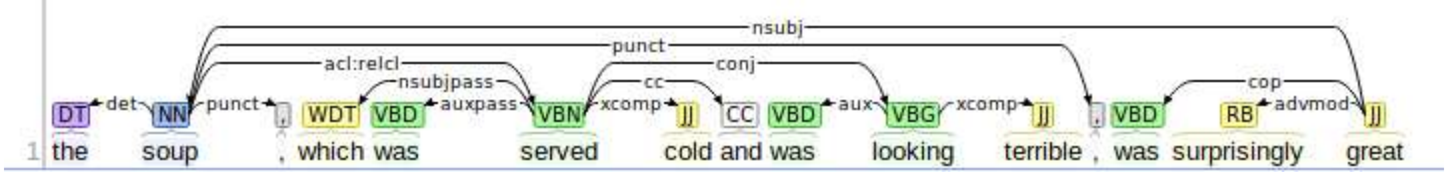
Real world:

the <u>soup</u> that the server recommended <u>was</u> amazing.
the <u>soup</u>, which I expected to be good, <u>was</u> actually pretty bad
the <u>soup</u>, which <u>was</u> served cold and <u>was</u> looking terrible, <u>was</u> surprisingly great.

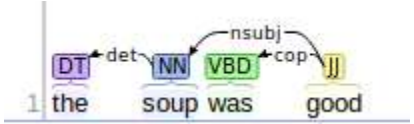
Example - review sentiment analysis

Syntax to the rescue!

Capture this:



With a single declarative pattern:

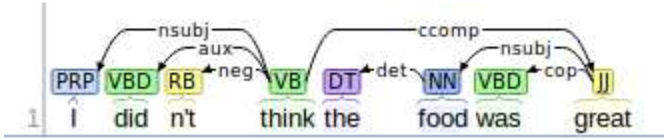


Can help to generalize and transfer knowledge across domains

Example - review sentiment analysis

Until it doesn't:

I didn't think the food was great:



Would combining rules with deep learning be the solution?

Thank you again!

Human: What do we want?
Computer: Natural-language processing!

Human: When do we want it?
Computer: When do we want what?

enliteAI

Dynamic Word Embeddings

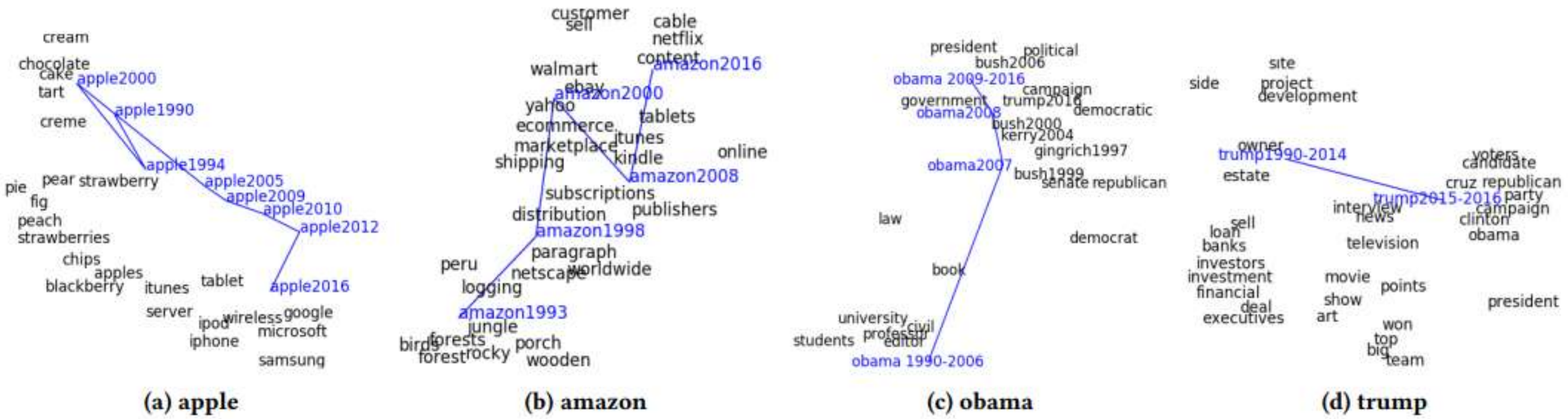


Figure 1: Trajectories of brand names and people through time: apple, amazon, obama, and trump.

Dynamic Word Embeddings for Evolving Semantic Discovery
https://arxiv.org/pdf/1703.00607.pdf