Natural Language Processing with Deep Learning

NGUYEN MINH TIEN
AI Researcher
AI Lab, Cinnamon
Hung Yen University of Technology and Education
CTM Palace, Hanoi

December 13, 2018

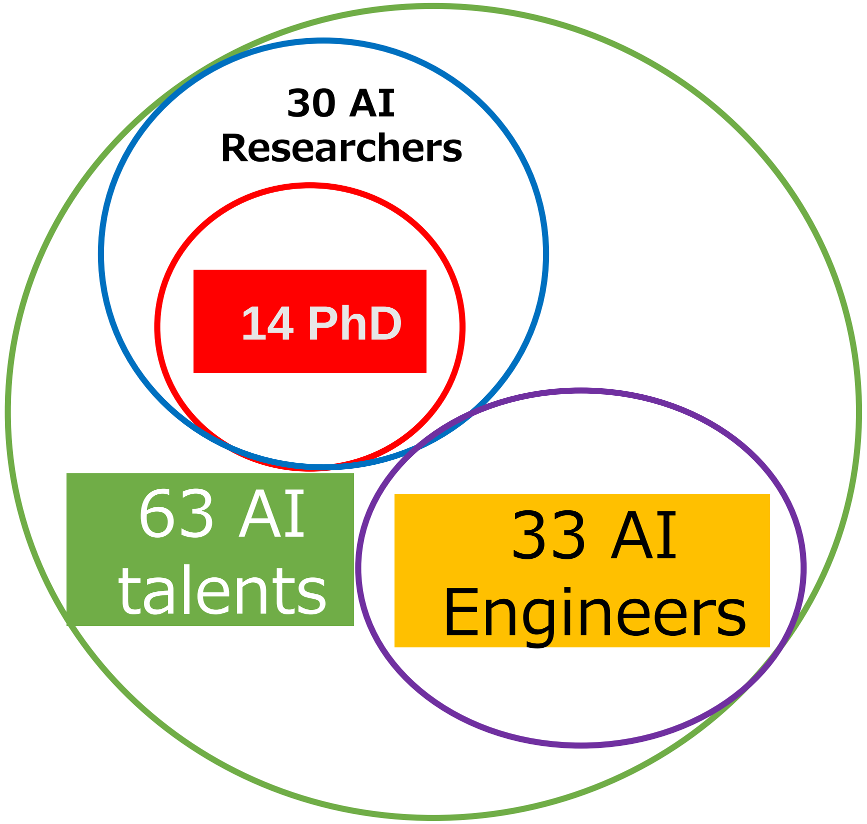
Who Am I

- 2012-2014: Master student, UET, VNUH
- 2013-2014: Visiting Researcher, National Institute of Informatics (NII), Japan
- 2015-2018: PhD candidate, Japan Advanced Institute of Science and Technology (JAIST)
- Site: https://sites.google.com/site/minhtienhy/
- Google scholar:

Minh-Tien Nguyen
Japan Advanced Institute of Science and Technology (JAIST)
Verified email at jaist.ac.jp
Machine Learning Deep Learning Natural Language Processing Text Summarization

Cinnamon

Cinnamon AI Lab



- Top Japanese corporations trust Cinnamon
- 40+ Paying customers, 70+ in Sales

Top 10 banking	Top 20 insurance	Top 10 systems integrator

By your side, for life
DAI-ICHI LIFE
Dai-ichi Life Group

UNISYS

JCB

TIS
TIS INTEC Group

CTC
Challenging Tomorrow's Changes

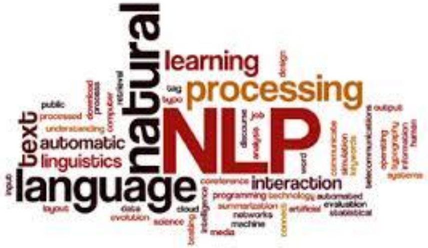
TOKIO MARINE NICHIDO
東京海上日動

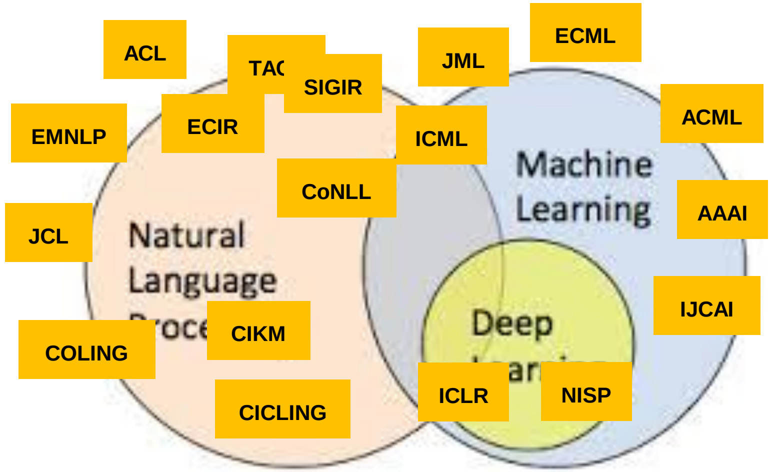
Content

- Introduction
- Natural Language Processing Problems
- Main Deep Learning Approaches in NLP
- State-of-the-art Achievements in NLP Tasks
- NLP Projects in Cinnamon

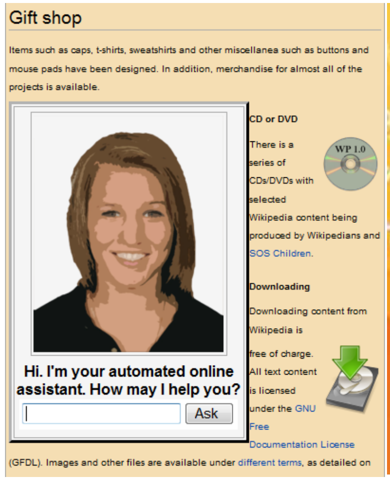
Introduction to NLP

- Understanding natural languages
- Sequence of text
- Using ML/DL as a tool for addressing NLP problems

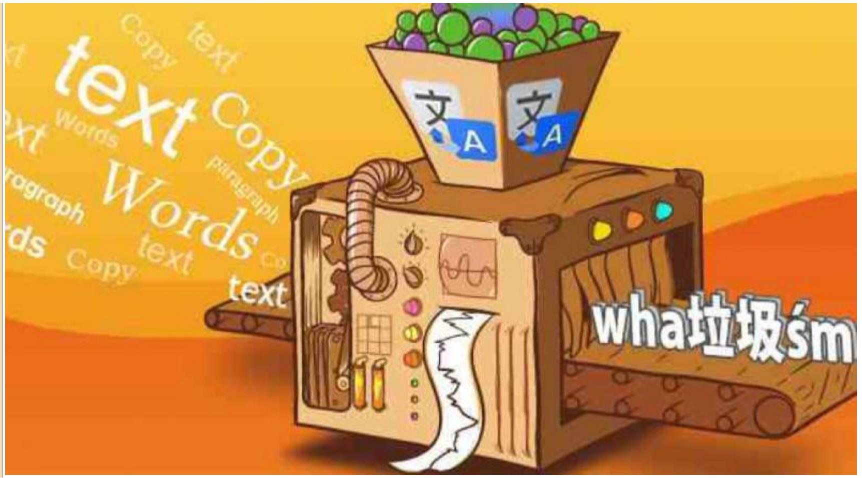




NLP and main venues



Automated online assistant



Machine translation

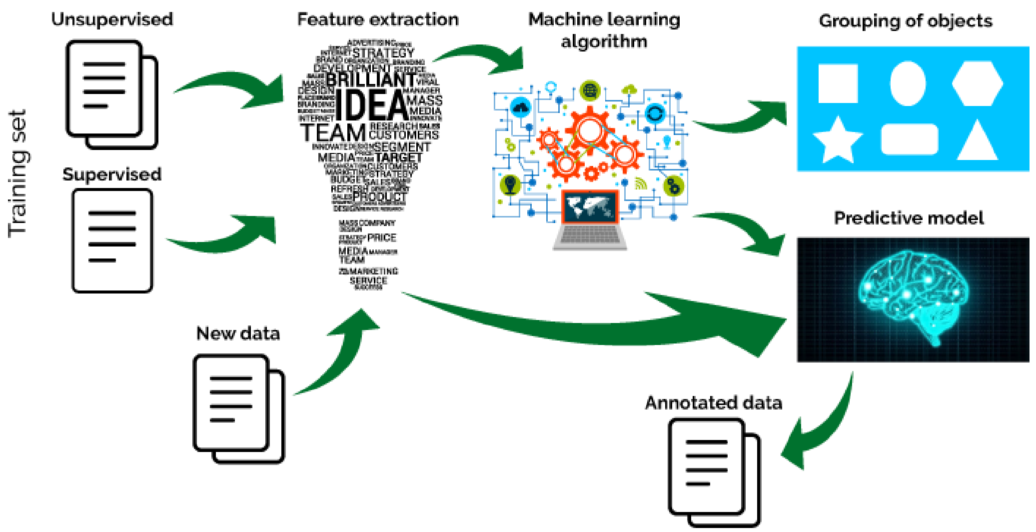
NLP with Machine Learning/Deep Learning

More Deeper Application of NLP

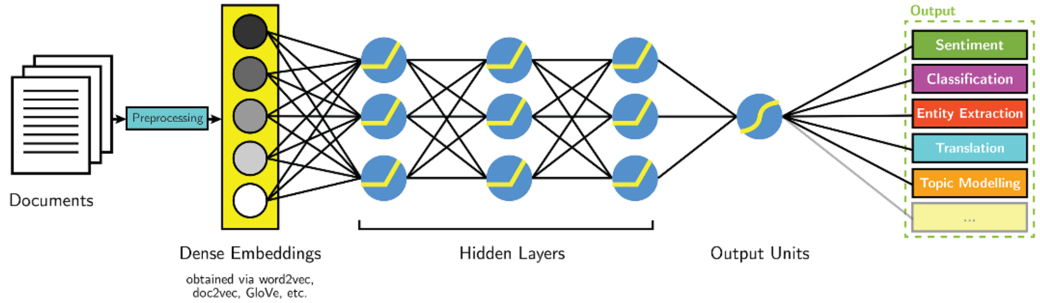
Group 1	Group 2	Group 3
Cleanup, Tokenization	Information Retrieval and Extraction (IR)	Machine Translation
Stemming	Relationship Extraction	Automatic Summarization/ Paraphracing
Lemmatization	Named Entity Recognation (NER)	Natural Language Generation
Part of Speech Tagging	Sentiment Analysis/Sentance Boundary Dismbiguation	Reasoning over Knowledge Based
Query Expansion	World sense and Dismbiguation	Quation Answering System
Parsing	Text Similarity	Dialog System
Topic Segmentationand Recognation	Coreference Resolution	Image Captioning & other Multimodel Tasks
Morphological Degmentation (Word/Sentences)	Discourse Analysis	

From core-NLP tasks to applications

Machine Learning



Deep Learning-based NLP



Pioneers of Deep Learning

Geoffrey Hinton
University of Toronto
Google Brain

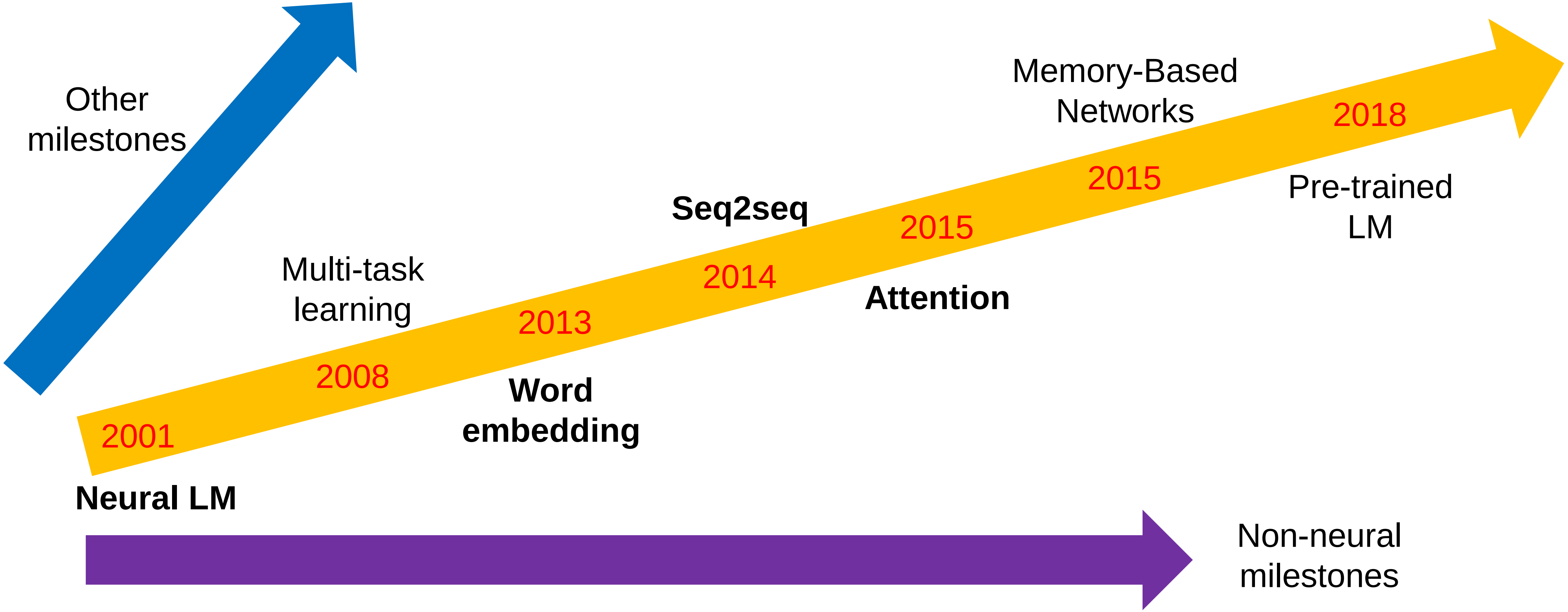
Yann LeCun
Director of AI Research, Facebook
Director of the NYU Center of Data Science

Yoshua Bengio

Prof. at Montreal University

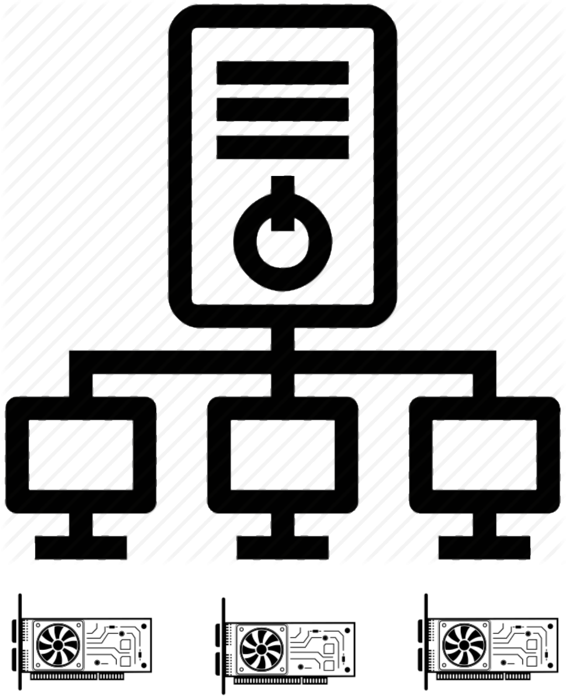
Andrew NG
Co-Chairman and Co-Founder of Coursera
Prof. at Stanford University

The Neural History of NLP



Question

Why DL is successful in many problems and applications?



BOW vs. Neural Representation

~100k columns

	good	movie	very	a	did	like
very	0	0	1	0	0	0
good	1	0	0	0	0	0
movie	0	1	0	0	0	0
very good movie	1	1	1	0	0	0

Spare vs. dense

~300 columns

	x_1	x_2	x_3
very	**0.7**	**0.4**	**0.5**
good	**0.2**	**-0.1**	**0.1**
movie	**-0.5**	**0.4**	**0.1**

Example: word2vec embeddings

~300 columns

	x_1	x_2	x_3
very	**0.7**	**0.4**	**0.5**
		+	
good	**0.2**	**-0.1**	**0.1**
		+	
movie	**-0.5**	**0.4**	**0.1**
		=	
very good movie	**0.4**	**0.7**	**0.7**

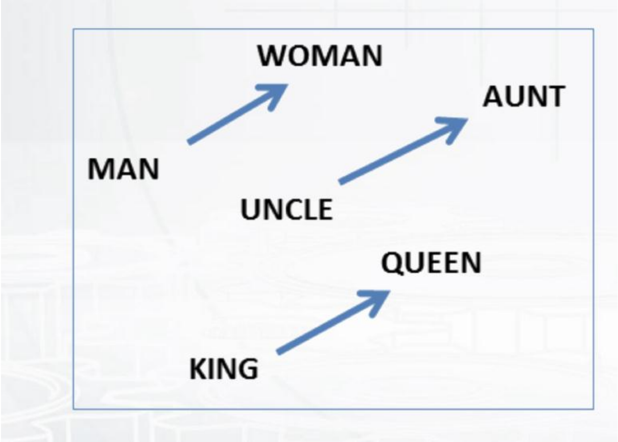
Example: word2vec embeddings

Neural representation: better for mapping text into the vector space model

Word Embedding

- **Word similarity**
- **Text representation by a dense model**

word relatedness



Two architectures:

- CBOW (Continuous Bag-of-words):

$$p(w_i | w_{i-h}, \ldots w_{i+h})$$

I love to play football

- Continuous Skip-gram:

$$p(w_{i-h}, \ldots w_{i+h} | w_i)$$

I love to play football

vector('king') - vector('man') + vector('woman') ~ vector('queen')

Word2vec:
- Open source: https://code.google.com/archive/p/word2vec/
- No need labeled data
- Can apply for any language

toolkit

Language Model

Why do we need LM

- Suggestions in text
- Spelling correction
- Machine translation
- Speech recognition
- Handwriting recognition
- N-grams model, N = 1, 2, 3⋯

$$P(w_t|\text{context})\ \forall t \in V.$$

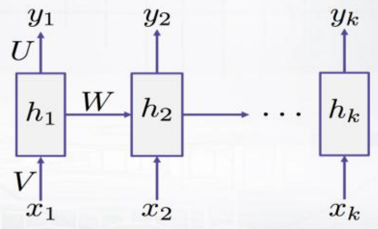
Counting way (classical)

This is the house that Jack built.
This is the malt
That lay in the house that Jack built.
This is the rat,
That ate the malt
That lay in the house that Jack built.
This is the cat,
That killed the rat,
That ate the malt
That lay in the house that Jack built.

$$p(house \mid this\ is\ the) = \frac{c(this\ is\ the\ house)}{c(this\ is\ the\ ...)} = \frac{1}{4}$$

- Extremely popular architecture for any sequential data:

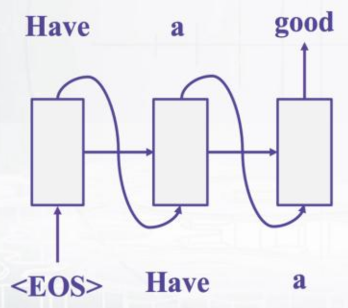
$$h_i = f(Wh_{i-1} + Vx_i + b)$$
$$y_i = Uh_i + \tilde{b}$$

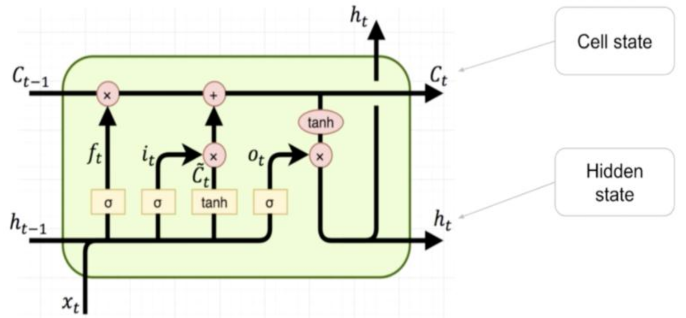


LM with LSTM

Idea:

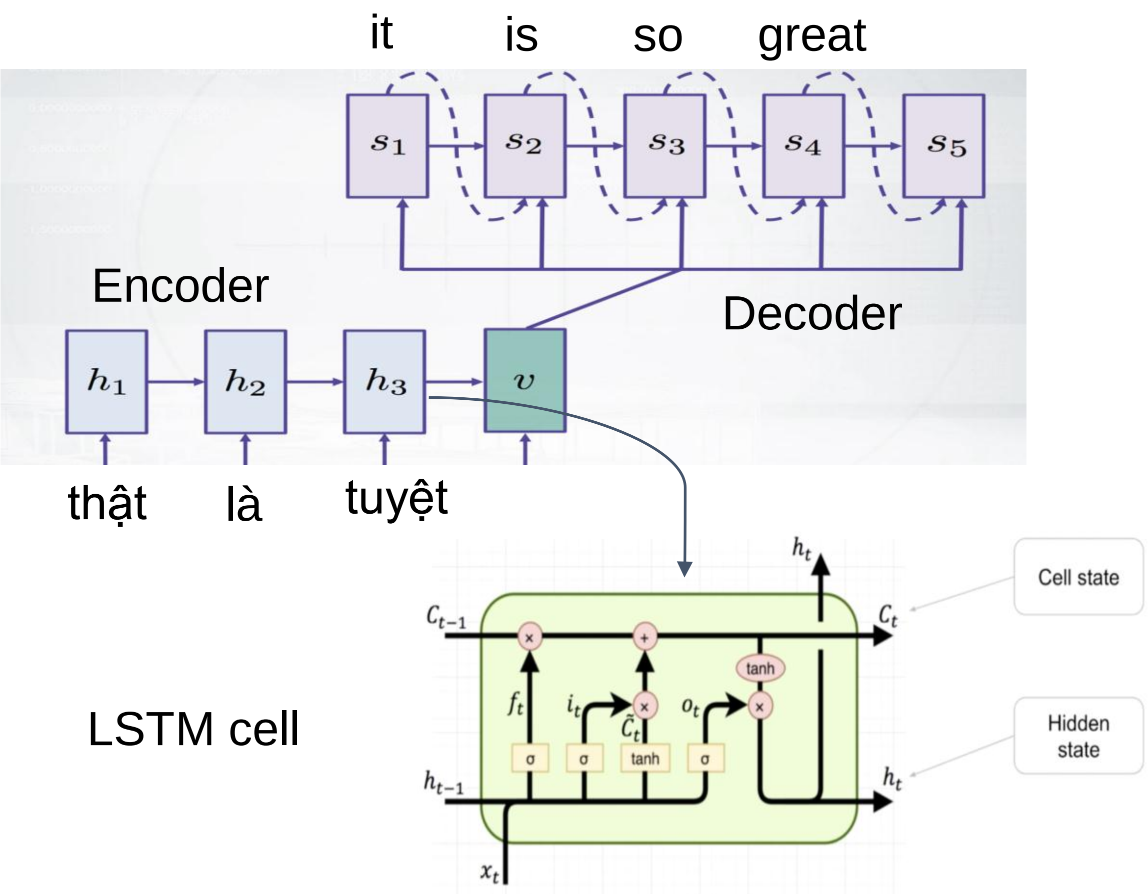
- Feed the previous output as the next input
- Take *argmax* at each step (greedily) or use *beam search*





LSTM Cell

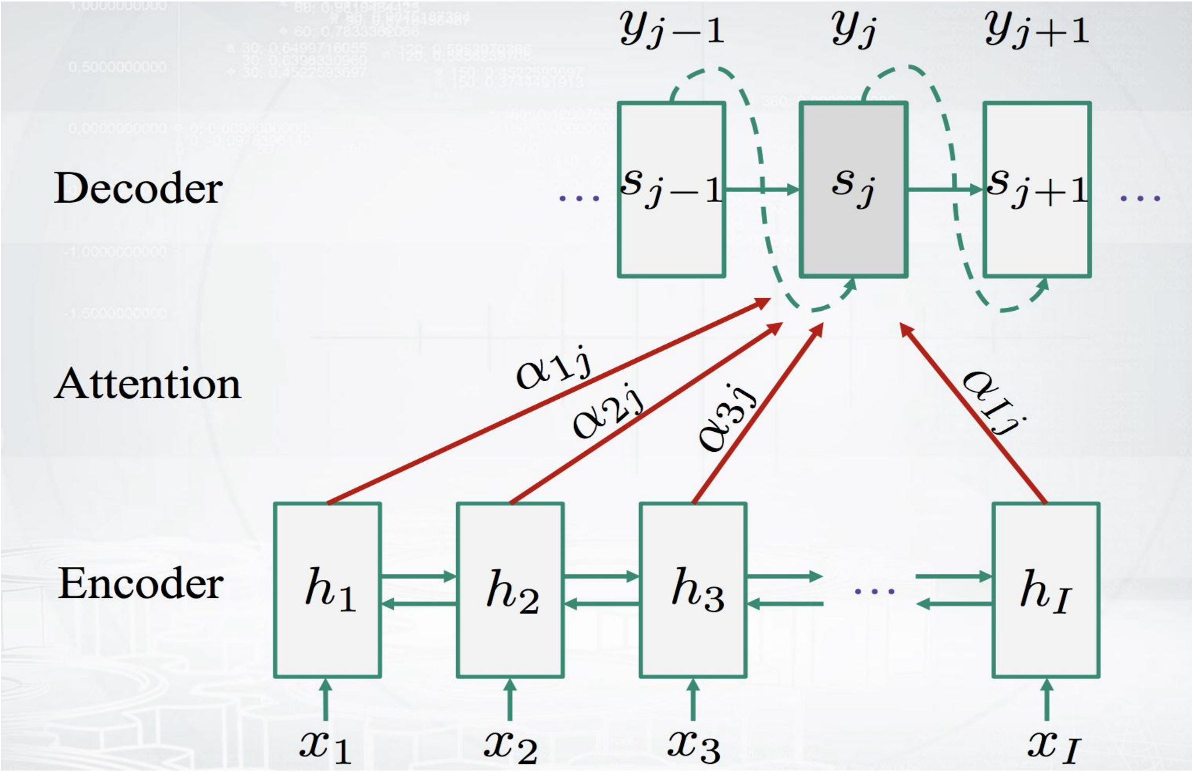
Neural Machine Translation



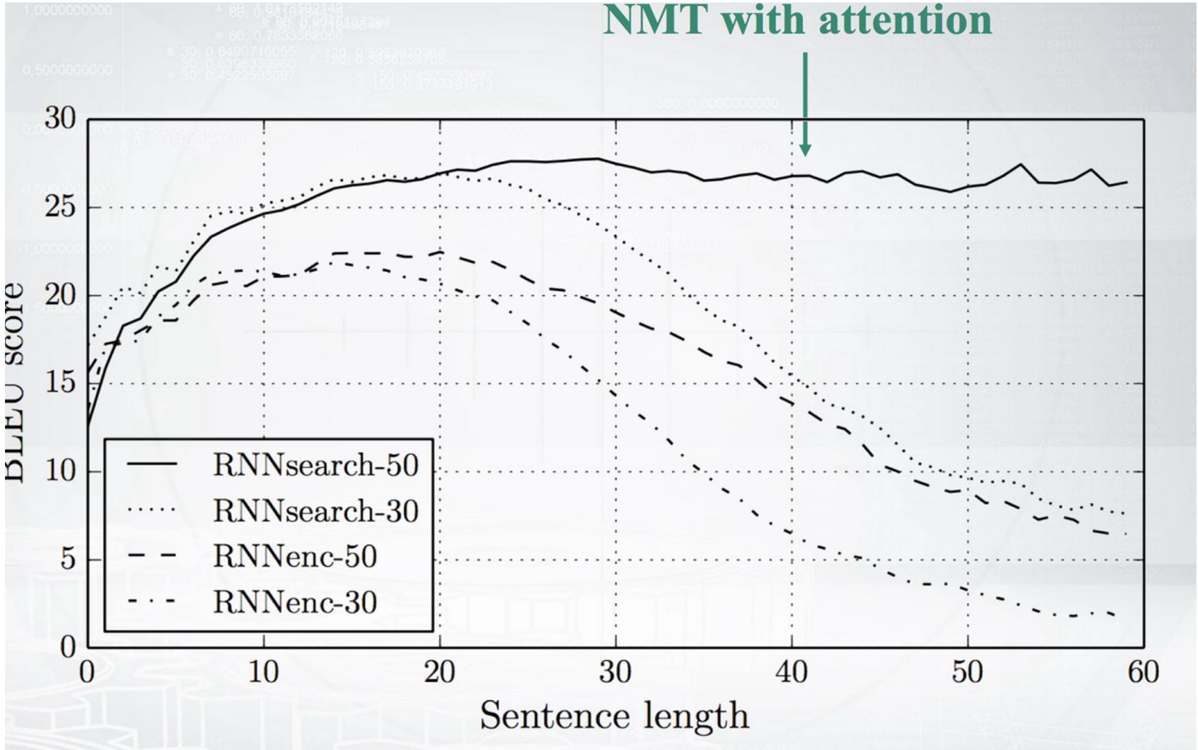
- Translating a sequence in source language, e.g. Vietnamese to a target language, e.g. English
- The final vector v is the input for the decoder
- We do not know which are important words in the decoding step
- Words in the input are treated as the same
- Need a mechanism for weighting words in the decoding step

Neural Machine Translation

Attention mechanism

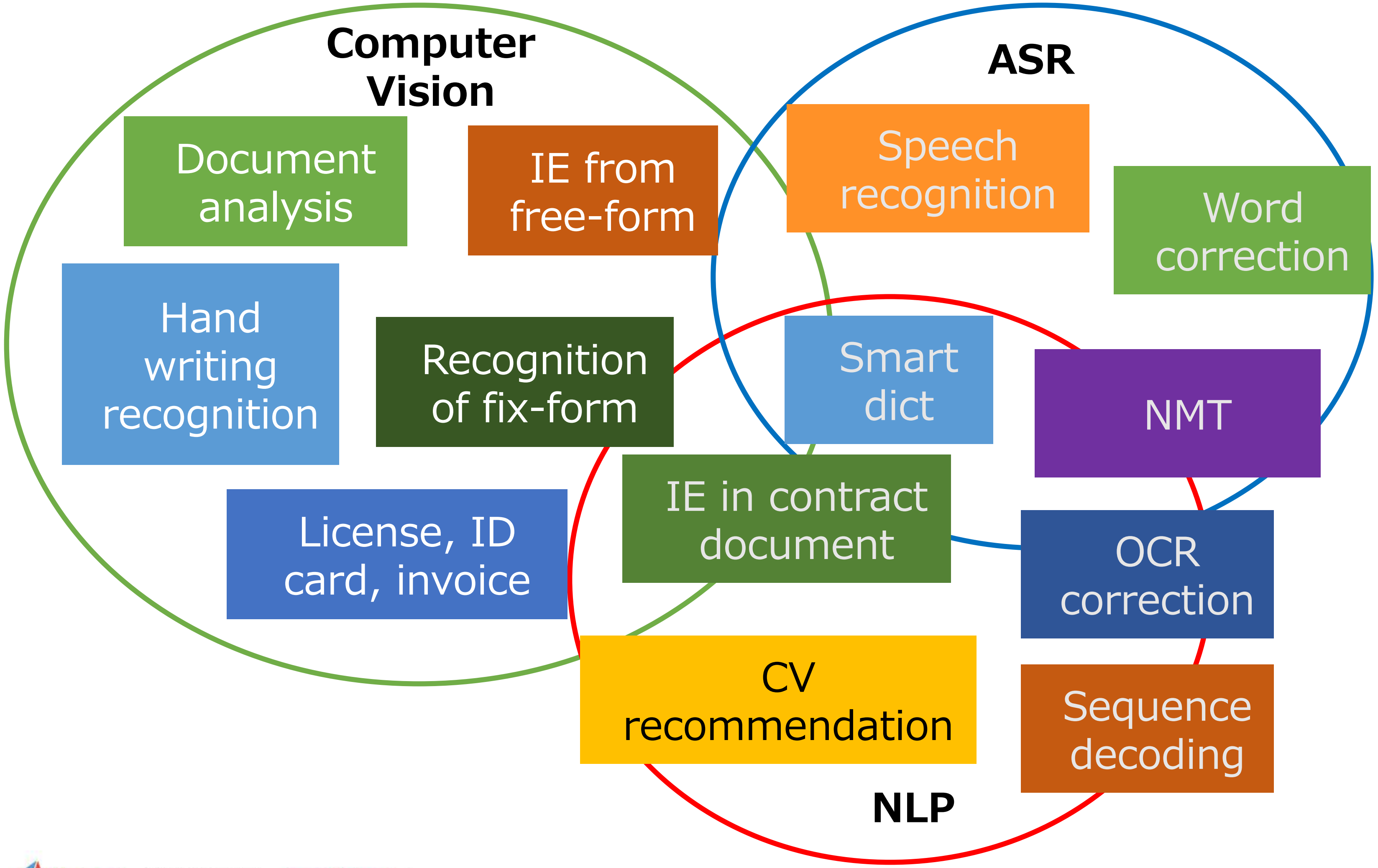


Results with long sentences



- Seq2Seq: final vector of encoder is input of decoder
- Attention: a weighted vector of input

Cinnamon's Research and Projects

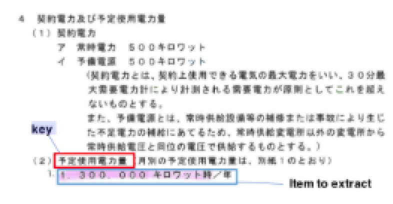
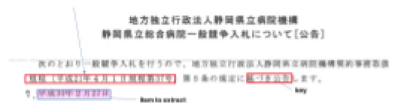
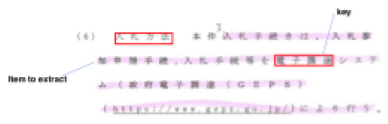
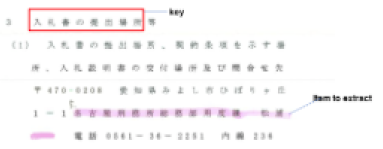


VIETNAM WEB SUMMIT

APPLANCER Hire Professionals, Get Great Apps

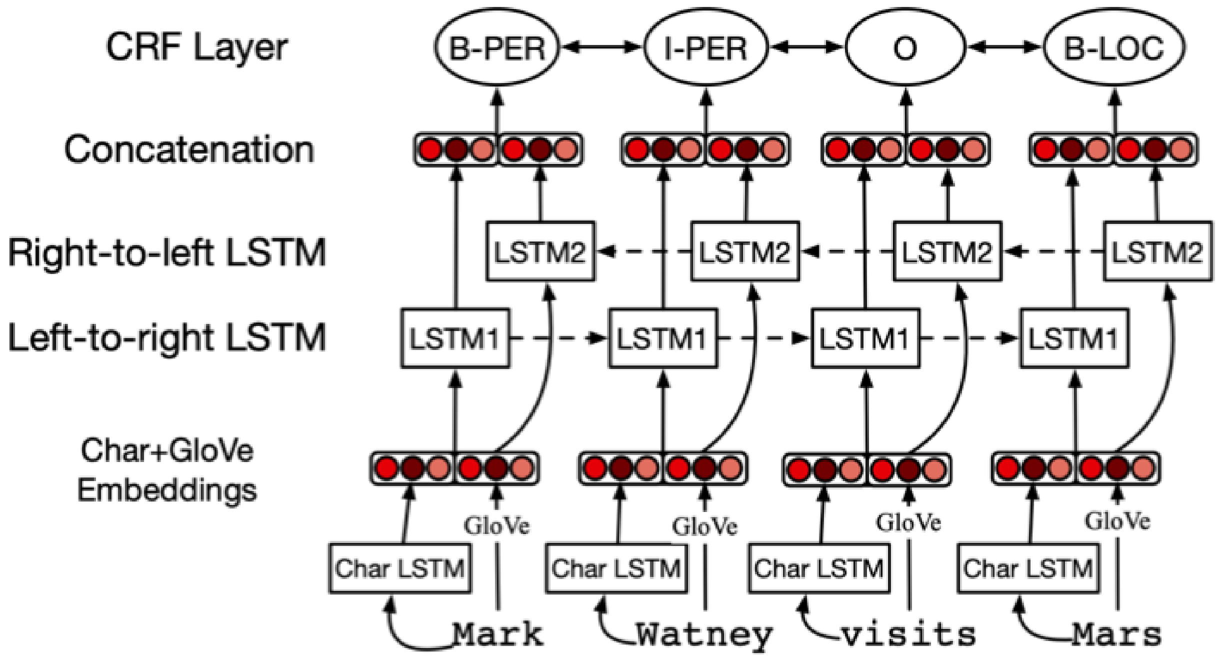
TOPDev

Information Extraction in Contract Document

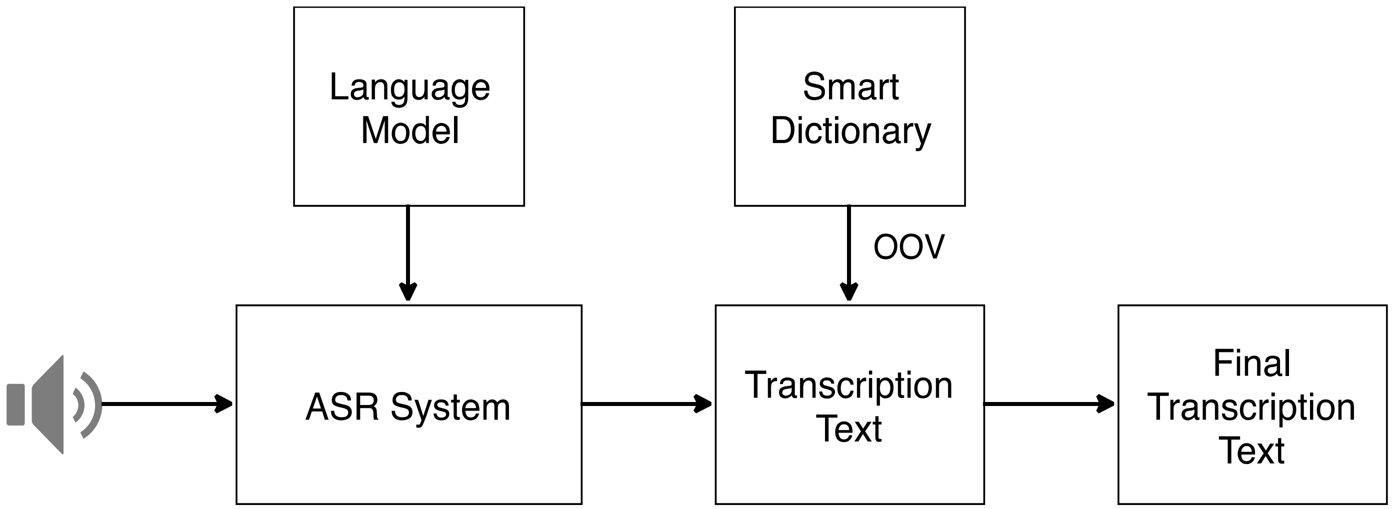
Tag	Meaning	Sample
電力量kWh	Total amount of electric power[kWh] for contract - number	
公告日	Public announcement date of bid - date	
入札方法	How to apply the bid - manner	
開札日	Opening date to bid - date	
資格申請送付先担当部署・名	Company name and department for submitting application of qualifications - name	
契約電力量kWh	Each electric energy amount(If there are some contracted spots)	

Extraction of tags

- NER Neural architecture
- NER by using a rule-based method
- Relation extraction



Smart Dictionary in ASR



Smart dictionary

- Dealing with OOV in ASR, e.g. "NRI"
- Word correction, e.g. "friend"
- Combined with NLM for correction

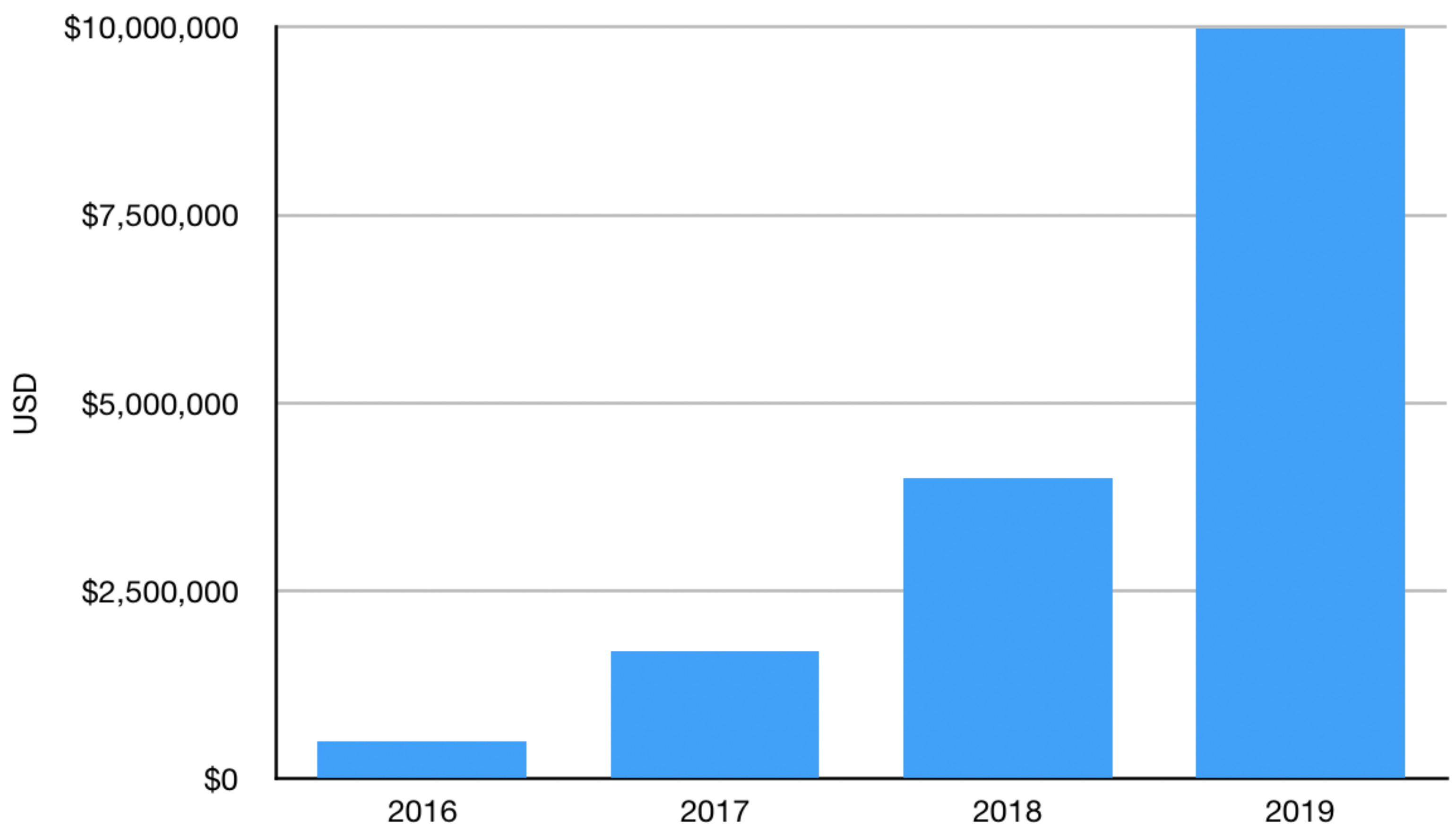
The first version

- A homonym Japanese dictionary
- A company EN-JP dictionary

Cinnamon knowledge base

- Smart dictionary
- Ontology
- WordNet
- Entities
- ....

Last but not least



The growth of Cinnamon

Goal: 1B in 2022

Join with us

- **AI Researcher**
- **AI Engineer**
- **Software Engineer**

Working on:

- Image Processing, NLP, Speech to Text Projects
- No limitation to **locations** or **ages**.

Contact us for details:
talent@cinnamon.is

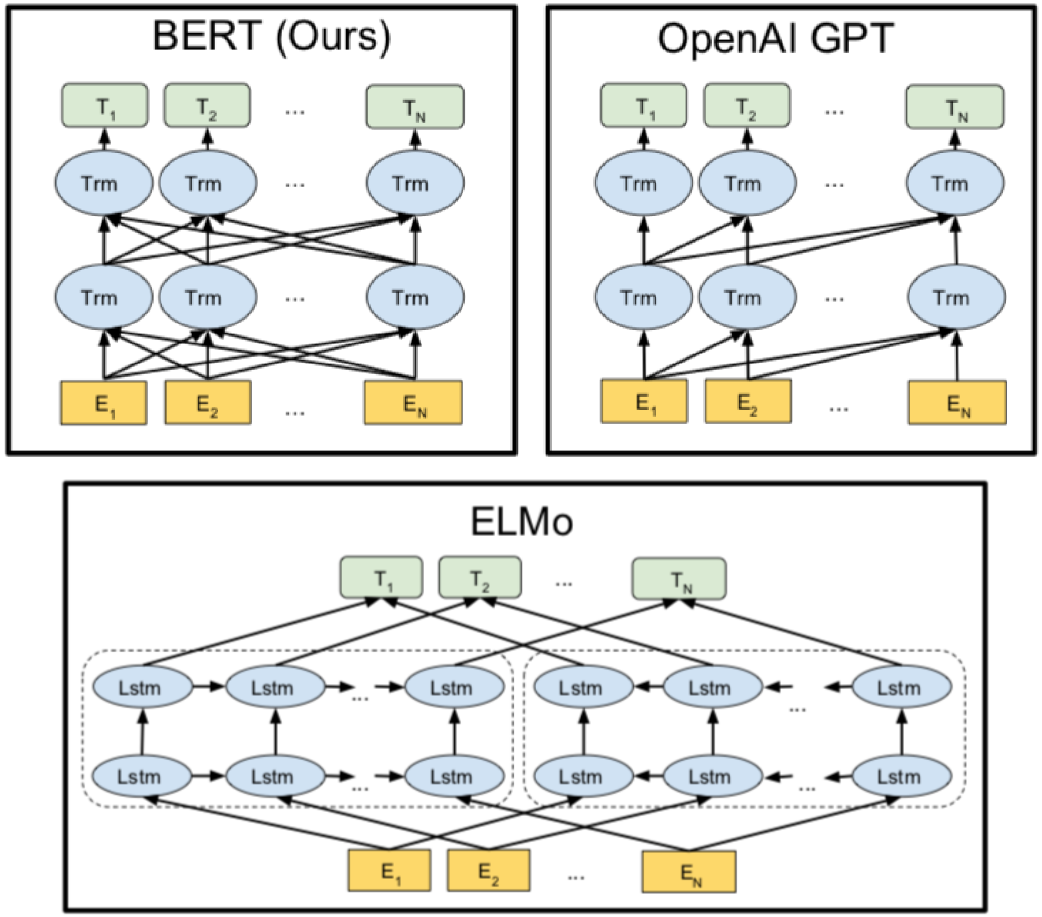
CV Recommendation System

Question & Answering

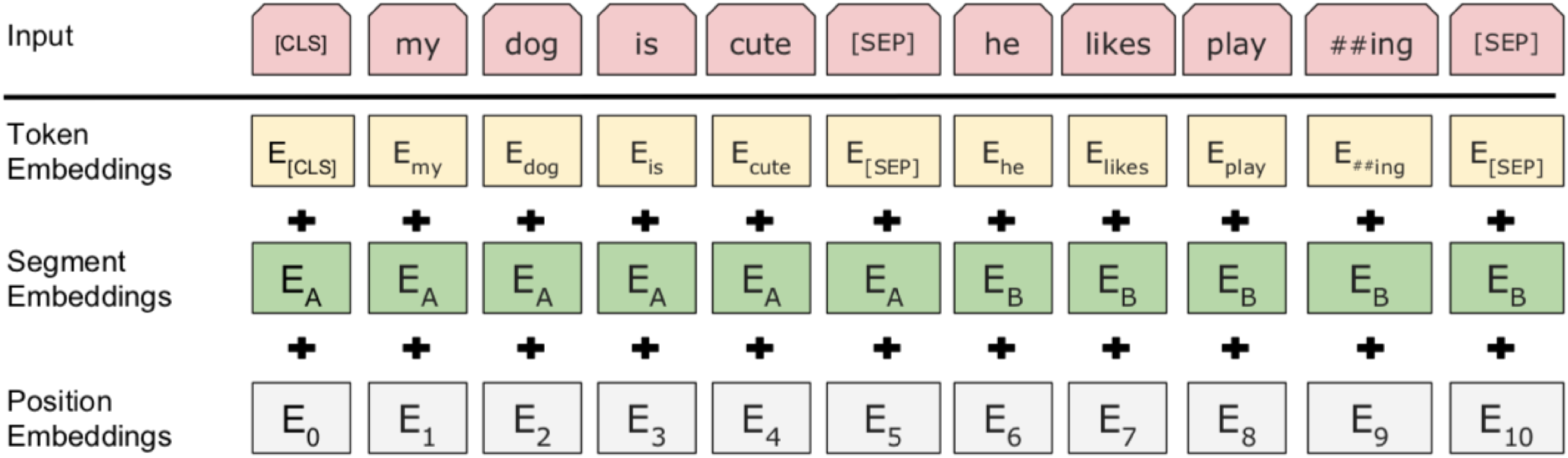
Thank you!

BERT

- Context-agnostic
- Only used to initialize the first layer of models
- Beneficial in many tasks
- **BERT**: Bidirectional Encoder Representations from Transformers
- Pre-train deep bidirectional representations by **jointly conditioning on both left and right context in all layers**
- BERT vs. OpenAI GPT vs. ELMo



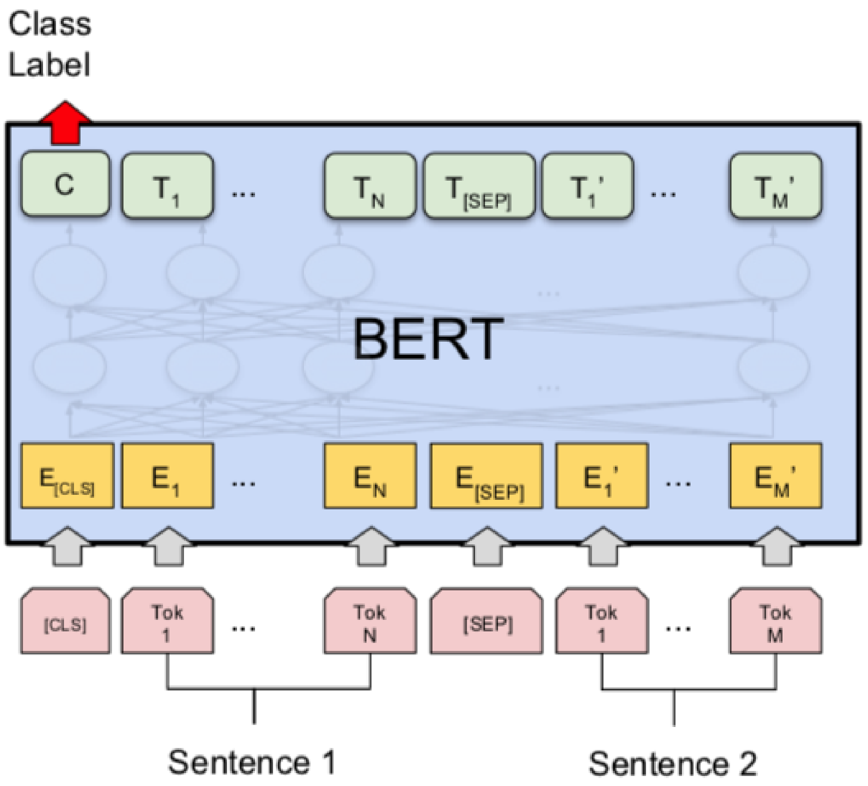
BERT Input Representation



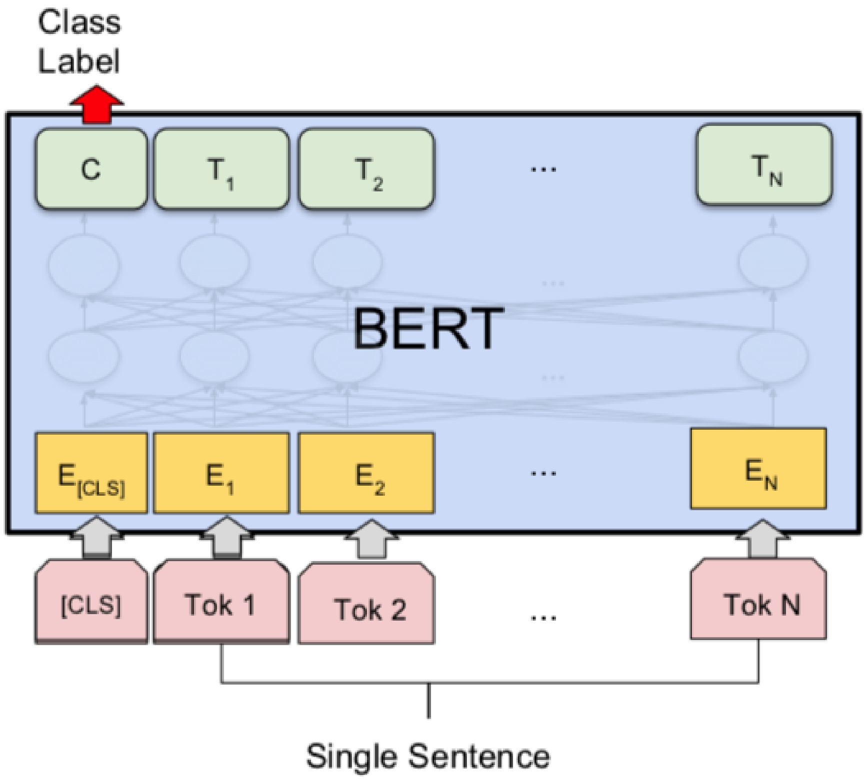
BERT input representation. The input embeddings is the sum of the token embeddings, the segmentation embeddings and the position embeddings.

Devlin et. al - BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018.

BERT for Challenging Tasks



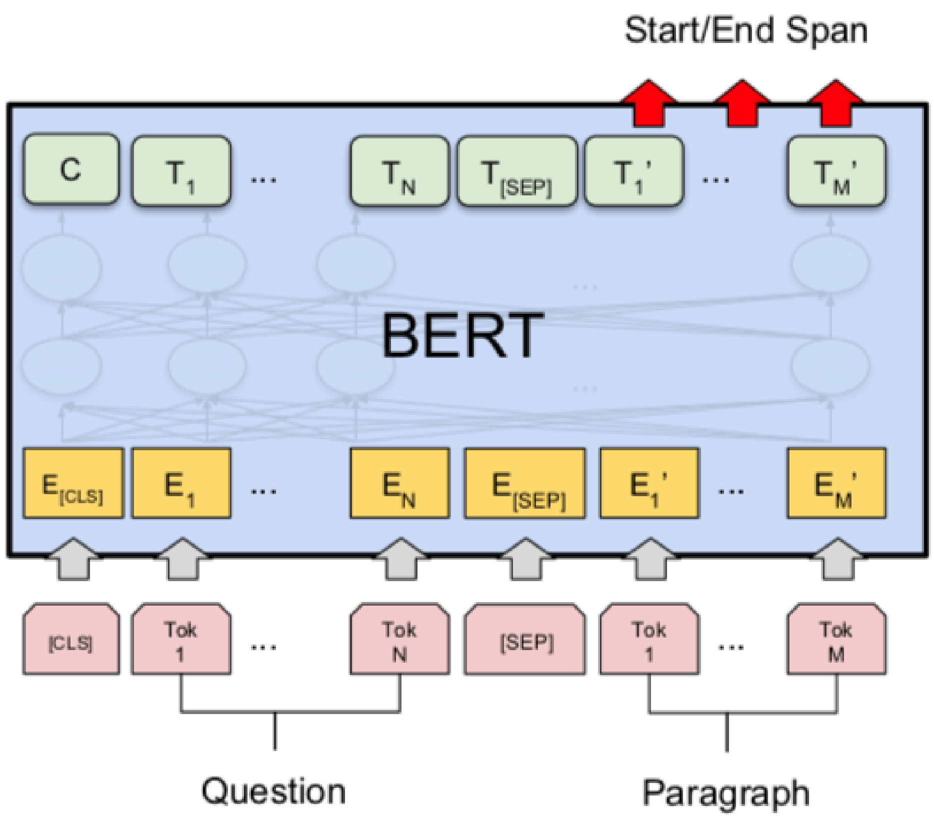
Sentence pair classification



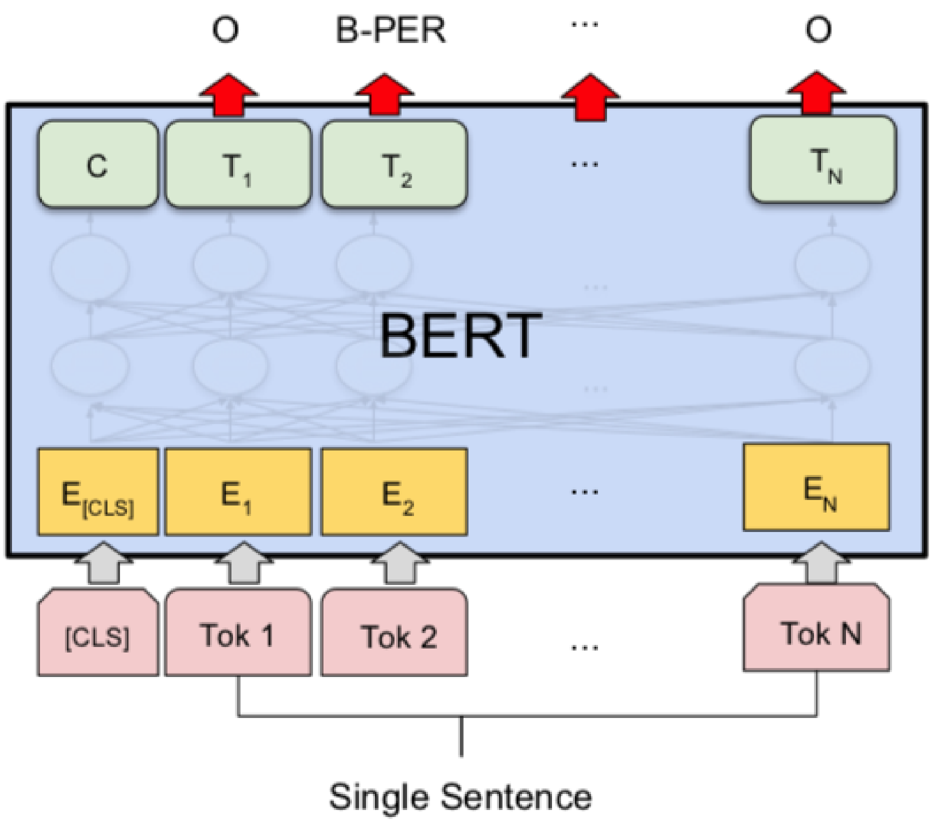
Sentence classification

Devlin et. al - BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018.

BERT for Challenging Tasks



Question answering task



Single sentence tagging

Devlin et. al - BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018.