Machine Translation

PV061

Pavel Rychlý

NLP Centre, FI MU

20 Sep 2023

Technical information

Technical information

- Pavel Rychlý
  - head of NLP Centre
- Natural Language Processing Centre
  - around 10 PhD students
  - you can be part of it (PV173 = 3 credits each semester)

Technical information

- Study materials in IS
  - book: Philipp Koehn: Neural Machine Translation (U366)
- Exam: written – max 10 questions
  - open books (offline)
  - max 60 points
- 30 point to pass
- extra points (max 30) for homeworks, projects
  - find good examples, illustrations to improve understanding
  - code, language, pictures
- exam, homeworks, … in English, Czech, Slovak

Previous knowledge

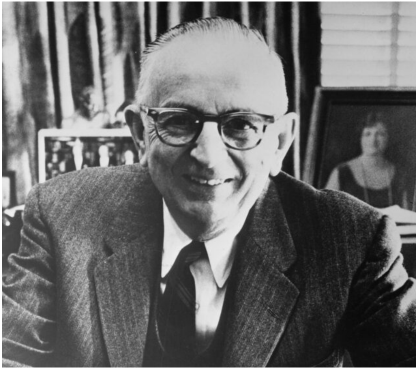
- no special requirements
  - reading mathematics
  - probabilities
- examples in Python
  - NumPy, PyTorch (matrix operations)
- complements
  - PV021: Neural Networks
  - PA153: Natural Language Processing
  - IA161: Natural Language Processing in Practice

History

Initial Idea

Warren Weaver on translation as code breaking (1947):

> *When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode".*



Translation or transcription

archæology office

ARCHAEOLOGY OFFICE

Translation or transcription

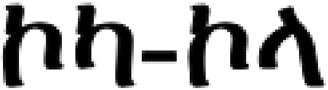
archæology office

ARCHAEOLOGY OFFICE

አማርኛ ለብዙ ዘመናት በእጅ ሲጻፍ ቆይቶ በ፲፱፻ ዓ.ም. ገደማ በማተሚያ መሣሪያ ሊከተብ ችሏል። የግዕዝን ፊደል በዶ/ር አበራ ሞላ ፈጠራ ወደ ኮምፕዩተር ገብቶ ከ፲፱፻፹ ዓ. ም. ወዲህ በኮምፕዩተር መጠቀም ከመቻሉም ሌላ በዩኒኮድ ዕውቅና አግኝቷል። ይህ የተስፋፋ ገጽም የቀረበው በእዚሁ ሥርዓት ሲሆን በቋንቋው የተጻፉ መረጃዎች ቊጥሮች እያደጉ ነው። የዓማርኛ ፊደልና ቋንቋም ዕውቅና እያደገ ስለመጣ በእጅ ስልኮችም በሚገባ

Translation or transcription

We need some examples



Translation or transcription

We need some examples

ኮካ-ኮላ

Coca-Cola

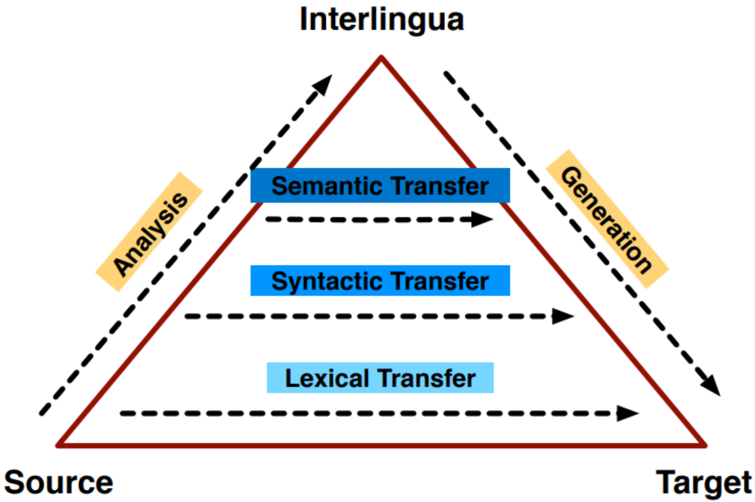
Early Efforts

- Excited research in 1950s and 1960s
- 1954 - Georgetown experiment

  Machine could translate 250 words and 6 grammar rules
- 1966 ALPAC report:
  - only $20 million spent on translation in the US per year
  - no point in machine translation

Main Idea

We can tranlate/transcribe on different levels



Rule-Based Systems

- Rule-based systems
  - build dictionaries
  - write transformation rules
  - refine, refine, refine
- Météo system for weather forecasts (1976)
- Systran (1968)

Statistical Machine Translation

- 1980s: IBM
- 1990s: increased research
- Mid 2000s: Phrase-Based MT (Moses, Google)
- Around 2010: commercial viability

Neural Machine Translation

- late 2000s: successful use of neural models for computer vision
- Since mid 2010s: neural network models for machine translation
- 2016: Neural machine translation the new state of the art

Results

- CUBBITT system for EN to CS
  - UFAL, Faculty of Mathematics and Physics, Charles University
- Nature Communications paper
  - September 2020
  - Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals
- better than human in adequacy in certain circumstances
  - news domain
  - rare phrases, translated literally by human translators

Handling problems

Word Translation Problems

- Words are ambiguous
- How do we find the right meaning, and thus translation?
- Context should be helpful

He deposited money in a **bank** account with a high **interest** rate.

Sitting on the **bank** of the Mississippi, a passing ship piqued his **interest**.

Syntactic Translation Problems

- Languages have different sentence structure
- Convert from object-verb-subject (OVS) to subject-verb-object (SVO)

das	behaupten	sie	wenigstens
this	claim	they	at least
the		she	

- Ambiguities can be resolved through syntactic analysis
  - the meaning **the** of **das** not possible (not a noun phrase)
  - the meaning **she** of **sie** not possible (subject-verb agreement)

Semantic Translation Problems

- Pronominal anaphora

  *I saw the movie and **it** is good.*

- How to translate **it** into German (or French)?
  - *it* refers to **movie**
  - **movie** translates to **Film**
  - **Film** has masculine gender
  - ergo: **it** must be translated into masculine pronoun *er*

Semantic Translation Problems

- Coreference

  Whenever I visit my uncle and his daughters,

  *I can't decide who is my favorite **cousin**.*

- How to translate cousin into German? Male or female?
- Complex inference required

Semantic Translation Problems

- Discourse

  Since *you brought it up, I do not agree with you.*

  Since *you brought it up, we have been working on it.*

- How to translated **since**? Temporal or conditional?
- Analysis of discourse structure – a hard problem

Rules

- hard to find
- many exceptions, exceptions in exceptions, …
- only suitable for cases without data

Statistics

- probabilities/rules learned from data
- linguistic knowledge about the structure of languages

  (SVO, VSO, .., ADJ+NN, NN+ADJ, ...)
- NLP tools (tokenizers, lemmatizers, taggers, ...)
- sparsity of data, long tail problem

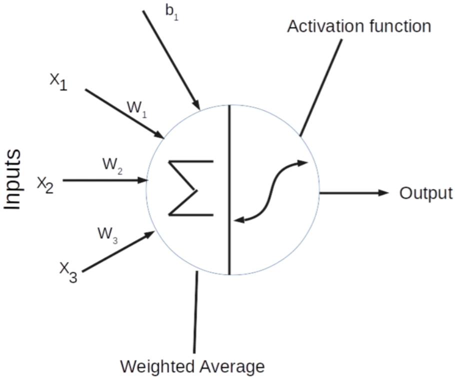
Neural Networks (NN)

- very simple model
    - neurons in layers
- trained on raw text data (almost no preprocessing)
- requires many training examples

Neural Networks

Neuron

- basic element of neural networks
- many inputs (numbers), weights (numbers)
- activation (transfer) function (threshold)
- one output: $y = \phi(\sum_{j=0}^{m} w_j x_j + b)$

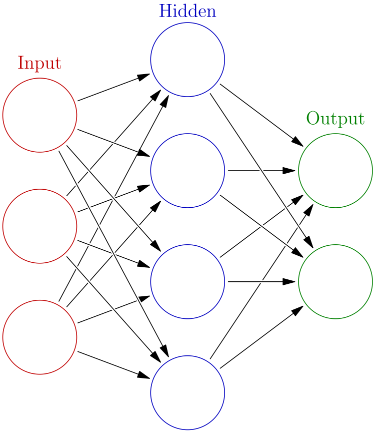


Neural Networks

- Input/Hidden/Output layer
- Input/output = vector of numbers
- hidden layer = matrix of parameters (numbers)

$$y_k = \phi(\sum_{j=0}^{m} w_{kj} x_j)$$

$$Y = \phi(WX^T)$$

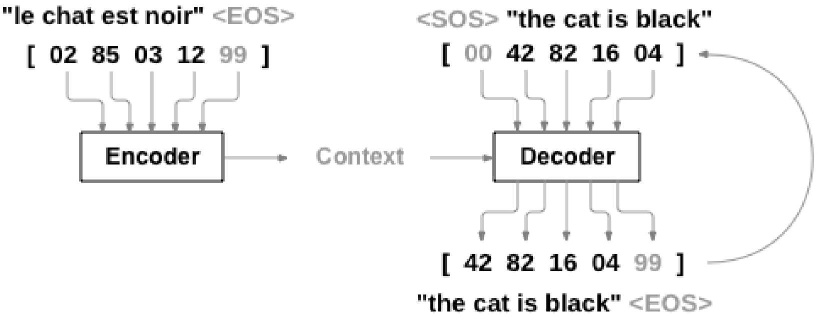


Words as vectors

$$continue = [0.286, 0.792, -0.177, -0.107, 0.109, -0.542, 0.349]$$

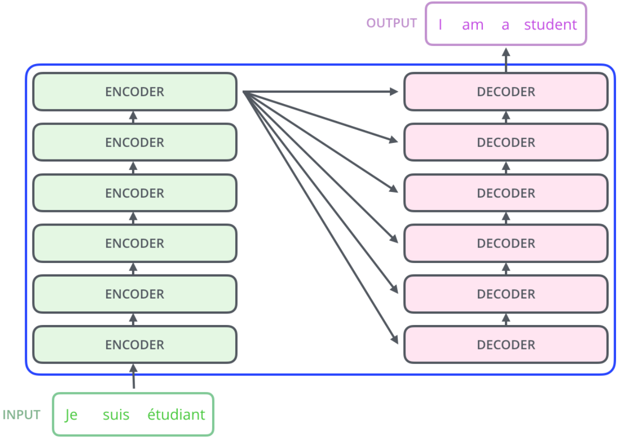
Neural Machine Translation

- encoder-decoder



Tranformer using attention

- each decoder layer has access to all hidden states from the last encoder
- use attention to extract important parts (vector)



Why are NNs better than statistics?

- continues space representation
  - words are not atomic
  - no sparsity problem
  - vectors handles relations
  - many realations, not explicit, unknown
- NN can represent any function (if deep enough)
  - structure of the function is not pre-defined

Why are NNs used only last 10 years?

- available big training data
- powerful hardware
  - matrix processing using specialized hardware GPU, TPU
- better learning strategies, NN optimizatons
- ready to use libraries/framewoks, datasets

Outline of the course

Outline 1

- statistics, probabilities
- language models
- IBM model 1
- phrase-base models
- decoding/generation
- evaluation

Outline 2

- neural networks, computation graphs
- tokenization, word representaion
- neural language models
- neural translation models
- monolingual data
- pretrained models

Conclusions

Current state

- MT systems use deep neural networks
- MT is very good in many areas

It can be improved

- more data
- bigger models
- better training strategies

MUNI
FACULTY
OF INFORMATICS