MACHINE LANGUAGE TRANSLATION

By
Surya Akella
Jigna Reshamwala

WHY MACHINE TRANSLATION?

6909 languages!!

CHALLENGES OF LANGUAGE TRANSLATION

- Not a word to word replacement
- Translation demands a deep understanding grammar and culture
- Different language structure
- Idioms and expressions
- Multiple Meanings of the same word

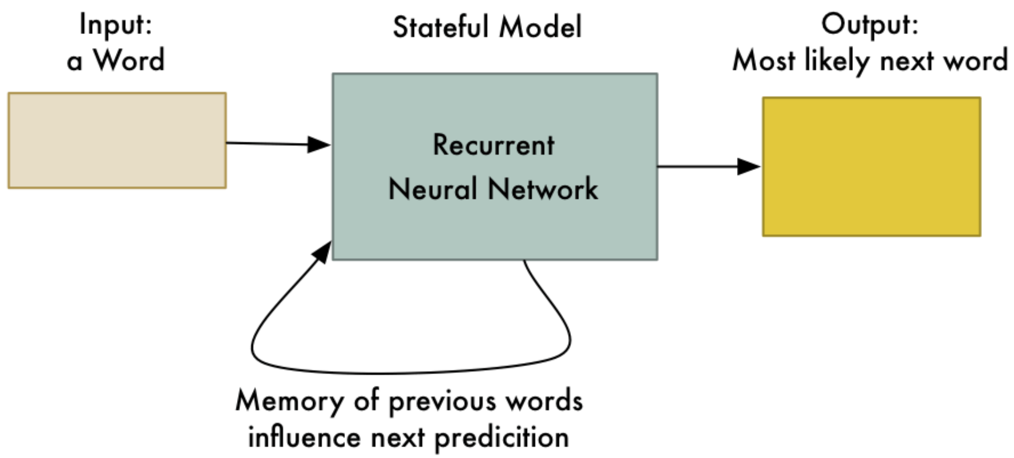
HISTORY OF MACHINE LANGUAGE TRANSLATION

- Rule Based, SMT and now NMT!!
- Google made the move from SMT to NMT last year. This reduced translation errors by an average of 60% compared to Google's phrase-based production system
- Google does Phrase Based Machine Translation but NMT can translate entire sentences at a time
- SMT still needed people to tweak the statistical models
- Challenges like word alignment and different word orders
- We need to develop a different model for every language pair, whereas NMT automatically learns patterns in the data using RNNs.

PROBLEM FORMULATION

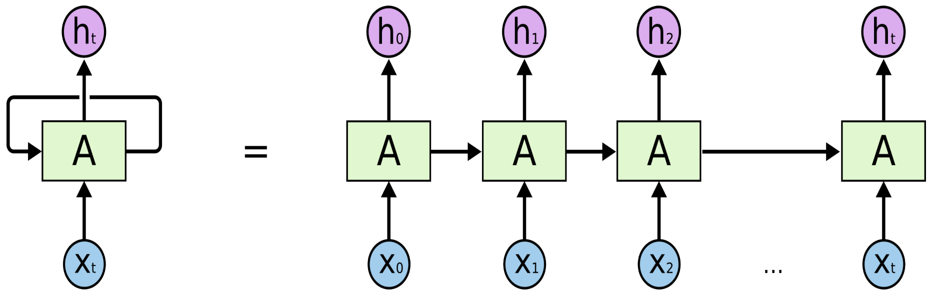
Recurrent Neural Network

Long Short-Term Memory (LSTM) units



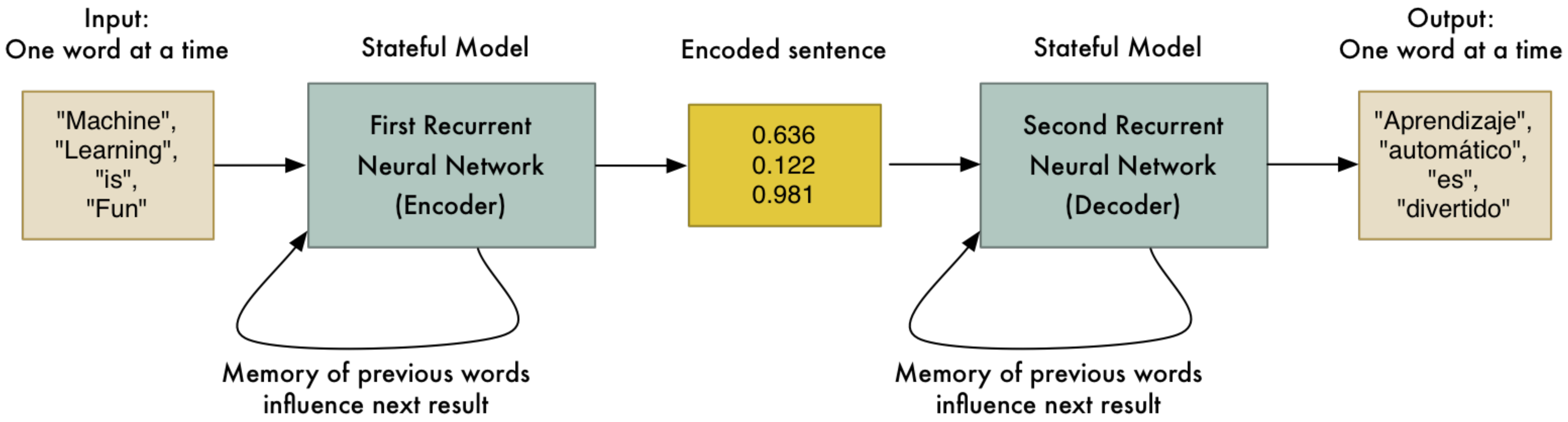
Output so far:

Machine



An unrolled recurrent neural network

SOLUTION



- Attention Mechanism
- Cost Function: Weighted cross-entropy loss for a sequence of logits
- Optimizer: RMSPropOptimizer
- Evaluation: BLEU Score

DATA DESCRIPTION

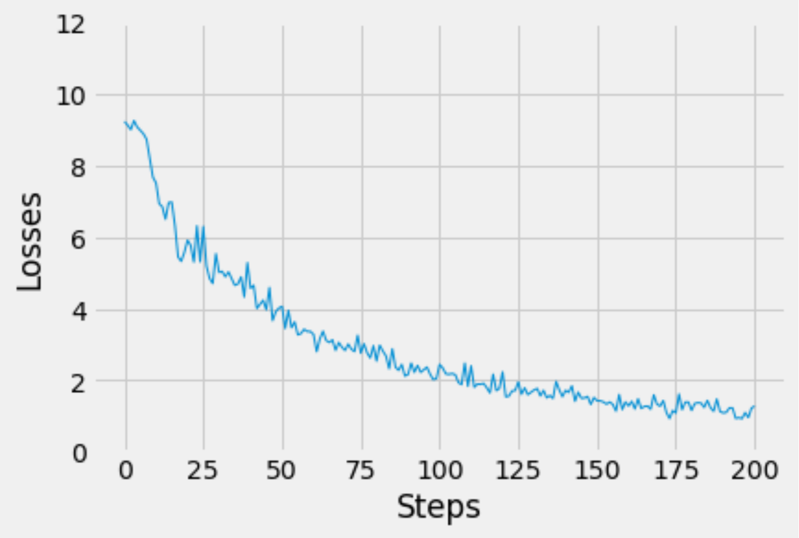
- Parallel corpus data, which is nothing but sentence aligned text in different languages
- Our Big Data: Europarl parallel corpus for Spanish & English, which contains text representation of European Parliament Proceedings[Around 371K sentences]
- Our Small Data: German to English [Around 38K sentences]
- Training Test Split: 90-10

RESULTS

- Works well with short and frequent sentences
- Poor results for rare words and phrases and longer sentences
- Poor results for rare words and phrases
- Final Avergae BLEU score 0.453899682606

```
1.
--------------------------------
What' s your name
Was ist du denn
--------------------------------
2.
--------------------------------
My name is
- Mein Name
--------------------------------
3.
--------------------------------
What are you doing
Was tust du da
--------------------------------
4.
--------------------------------
I am reading a book
Ich bin ein Buch Buch
--------------------------------
```

English to German



FUTURE WORK

- Newer Models and Architecture
- More data for training

THANK YOU!