TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Statistical Machine Translation

Statistical Machine Translation (SMT)

Phrase based SMT dominanted machine translation before deep Seq2Seq models.

SMT is still used for low resource languages, such as regional African languages, and in unsupervised machine translation.

Statistical Machine Translation (SMT)

Step I: Learn a phrase table — a set of triples (p, q, s) where

- p is a (short) sequence of source words.

- q is a (short) sequence of target words.

- s is a score.

("au", "to the", .5)   ("au banque", "for the bank", .01)

For a phrase triple P we will write P.source for the source phrase, P.target for the target phrase, and P.score for the score.

Derivations

Consider an input sentence x of length T.

We will write $x[s:t]$ for the substring $x[s], \ldots, x[t-1]$.

A derivation d from x is a sequence $(P_1, s_1, t_1,), \ldots, (P_K, s_K, t_K)$ where $P_k.\text{source} = x[s_k : t_k]$.

The substrings $x[s_k : t_k]$ should be disjoint and "cover" x.

For $d = [(P_1, s_1, t_1,), \ldots, (P_L, s_K, t_K)]$ we define

$$y(d) \equiv P_1.\text{target} \cdots P_K.\text{target}$$

We let $D(x)$ be the set of derivations from x.

Scoring

For $d \in D(x)$ we define a score $s(d)$

$$s(d) = \alpha \ln P_{\mathrm{LM}}(y(d)) + \beta \sum_k P_k.\mathrm{score} + \gamma \, \mathrm{distortion}(d)$$

where $P_{\mathrm{LM}}(y)$ is the probability assigned to string y under a language model for the target language

and distortion(d) is a measure of consistency of word ordering between source and target strings as defined by the indeces $(s_1, t_1), \ldots, (s_K, t_K)$.

Translation

$$y(x) = y(d^*(x))$$

$$d^*(x) = \operatorname*{argmax}_{d \in D(x)} \; s(d)$$

END