Alignment models for statistical translation

Thomas Debarre, Matthieu Kirchmeyer, Sylvain Truong, Nicolas Zhang

Introduction

This project explores three methods to solve the problem of word alignments for a bilingual corpus from the conventional mixtures models (IBM1, IBM2) to a first-order HMM model developped in [1].

The goal is to translate a text given in a language (French in all that follows) to another language (English) taking into account one-to-many alignments or null words; the quality of the translation relies on the quality of the word alignments.

We evaluated our algorithms on a small bilingual French-English dataset $\mathcal{D} = \{(\mathbf{f}^{(1)}, \mathbf{e}^{(1)}), \ldots, (\mathbf{f}^{(N)}, \mathbf{e}^{(N)})\}$ which consists of $N = 22$ French sentences composed of 4 to 10 words with their translated English equivalents.

Statistical translation models

Your text with scientific results or something...

$$H = \sum_{i=1}^{N} h_D(i) + \sum_{j>i=1}^{N} C_{ij} \tag{1}$$

In Ref. [1]... In Refs. [1, 2]... On webpage [3]...

IBM models

Table 1: This is a table with scientific results.

1	2	3	4	5
aaa	bbb	ccc	ddd	eee
aaaa	bbbb	cccc	dddd	eeee
aaaaa	bbbbb	ccccc	ddddd	eeeee
aaaaaa	bbbbbb	cccccc	dddddd	eeeeee
1.000	2.000	3.000	4.000	5.000

Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something...

HMM-based alignment models

The HMM-based alignment models were introduced by *Vogel and al.* and aim at modeling the joint probability of a French sentence $\mathbf{f}$, of length J and an alignment $\mathbf{a} \in \{1, \ldots, I\}^J$, given a English sentence $\mathbf{e}$ of length I. Without, loss of generality, this probability can be expressed as (chain-rule and independance of $\mathbf{f}$ and $\mathbf{a}$):

$$\begin{aligned} Pr(\mathbf{f}, \mathbf{a}|\mathbf{e}) &= Pr(f_0, a_0|\mathbf{e}) \prod_{j=1}^{J} Pr(f_j, a_j|f_1^{j-1}, a_1^{j-1}, \mathbf{e}) \\ &= Pr(f_0|\mathbf{e}) \cdot Pr(a_0|\mathbf{e}) \\ &\cdot \prod_{j=1}^{J} Pr(f_j|f_1^{j-1}, a_1^{j-1}, \mathbf{e}) Pr(a_j|f_1^{j-1}, a_1^{j-1}, \mathbf{e}) \end{aligned}$$

The HMM model reduced the generality of the previous formula by assuming first-order dependance of alignments and translation probability of French word at position j to be only dependent on the English word at position a_j, thus yielding the HMM model:

$$\begin{aligned} Pr(\mathbf{f}, \mathbf{a}|\mathbf{e}) &= p(f_0|e_{a_0}) p(a_0) \\ &\cdot \prod_{j=1}^{J} p(f_j|e_{a_j}) \cdot p(a_j|a_{j-1}, I) \end{aligned}$$

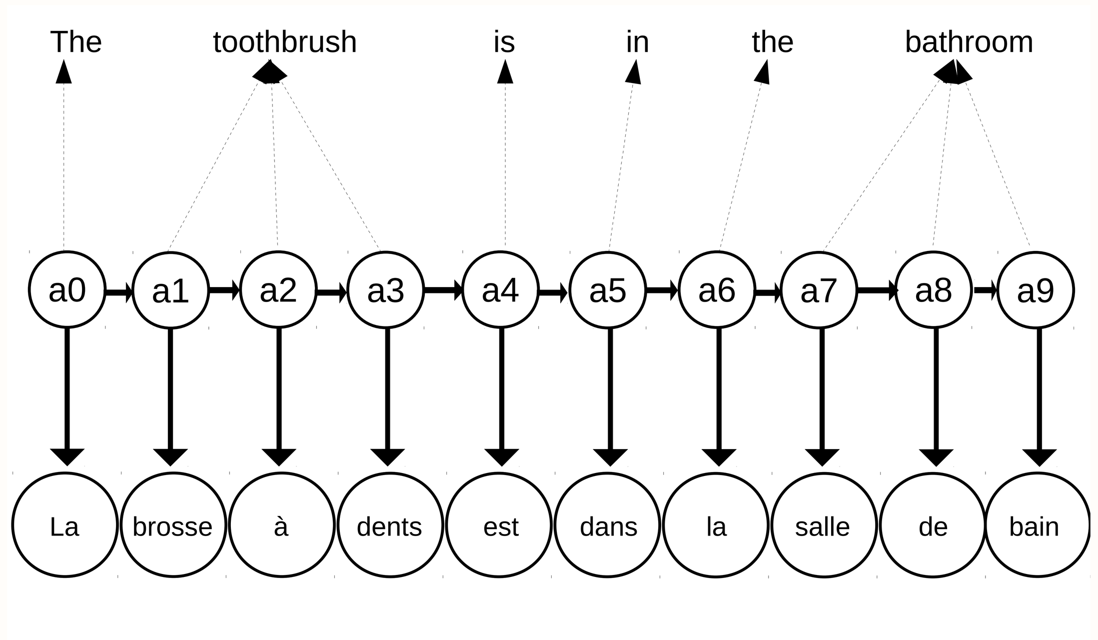

Figure 1: The HMM-based alignment model

Figure 1 presents the graphical model associated with HMM-based alignement models. What can be noticed is that, in this context, each sentence pair represents a Hidden Markov Model which parameters are shared with all other sentences of the corpus. The **parameters for the HMM model** are

$$\begin{aligned} \Theta = \{ & \\ p(i) & \quad \text{the initial alignment probabilities } (p(a_0)) \\ p(i|i', I) & \quad \text{the transition matrix} \\ p(f|e) & \quad \text{the emission (translation) probabilities}\} \end{aligned}$$

A further assumption from *Vogel and al.* is that alignment transition probabilities **only depend on relative positions**: $p(i|i', I) = p(i' - i|I)$

Just as for the IBM models, the observations are the sequences of French words in French sentences, while the alignments (and thus the corresponding English words) act as hidden variables. A way to solve this problem is to use an **EM principle**. The **Expectation step** consists in computing alignment unary and binary probabilities in each sentence pair $(\mathbf{f}, \mathbf{e})$.

$$\begin{aligned} \gamma_i(j) &= p(a_j = i|\mathbf{f}) \\ \xi_{k,l}(j) &= p(a_j = k, a_{j+1} = l|\mathbf{f}) \end{aligned}$$

This can be performed using the Forward-Backward algorithm. The **Maximisation step** consists in re-estimating the model's parameters according to the previously computed probabilities.

The **Viterbi** algorithm is then used to retrieve the most likely alignements in the corpus, according to the computed transition and emission probabilities.

Results and discussions

The experimental setup

Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something...

Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something...

Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something...

Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something...

Your text with scientific results or something... Your text with scientific results or something... Your text with scientific results or something...

Summary and conclusions

References

[1] S. Vogel, H. Ney, and C. Tillmann, HMM-based word alignment in statistical translation

[2] J. Doe, J. Smith, Other article name, *Phys. Rev. Lett.*

[3] http://www.google.pl