A STATISTICAL APPROACH TO MACHINE TRANSLATION

Author : **Peter F. Brown**
Professor : 陳嘉平
Repoter : 楊治鏞

Outline

- Introduction
- The language model
- The translation model
- Searching
- Parameter estimation
- Two pilot experiments
- Plans

Introduction(1/3)

- The field of machine translation is almost as old as the modern digital computer.

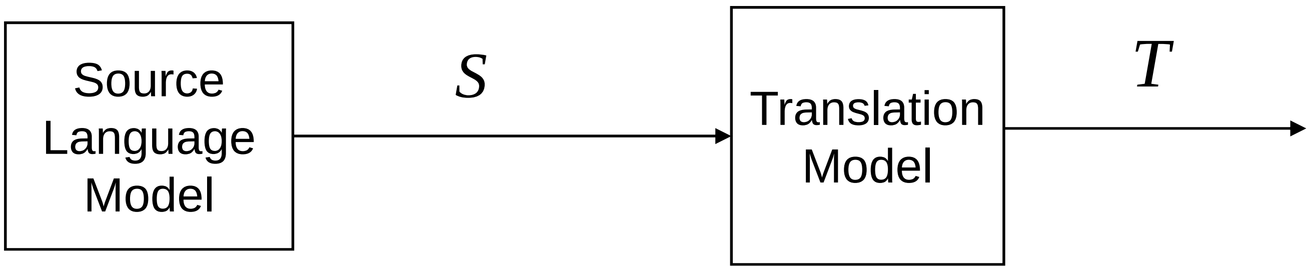
- We feel that it is time to give statistical methods a chance in machine translation.

Introduction(2/3)

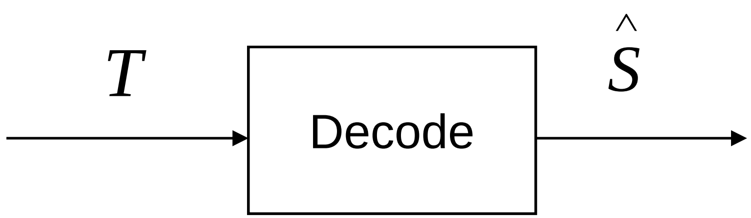
- We know that our chance of error is minimized by choosing that sentence S that is most probable given T. Thus, we wish to choose S so as to maximize $\Pr(S \mid T)$. Using Bayes' theorem, we can write

$$\Pr(S \mid T) = \frac{\Pr(S)\Pr(T \mid S)}{\Pr(T)}$$

Introduction(3/3)



$$\Pr(S) \qquad \times \qquad \Pr(T \mid S) = \Pr(S, T)$$



$$\hat{S} = \arg\max_{S} \Pr(S \mid T) = \arg\max_{S} \Pr(S, T)$$

The language model

- Given a word string, $s_1 s_2 \ldots s_n$

$$\Pr(s_1 s_2 \ldots s_n)$$

$$= \Pr(s_1)\Pr(s_2 \mid s_1)\ldots\Pr(s_n \mid s_1 s_2 \ldots s_{n-1})$$

- n-gram model
  - bigram
  - trigram
- We can see the power of a trigram model by applying it to something that we call bag translation from English into English.

Bag translation

- In bag translation we take a sentence, cut it up into words, place the words in a bag, and then try to recover the sentence given the bag.
- *Exact reconstruction* (24 of 38)

  Please give me your response as soon as possible.

  Please give me your response as soon as possible.
- *Reconstruction preserving meaning* (8 of 38)

  Now let me mention some of the disadvantages.

  Let me mention some of the disadvantages now.
- *Garbage reconstruction* (6 of 38)

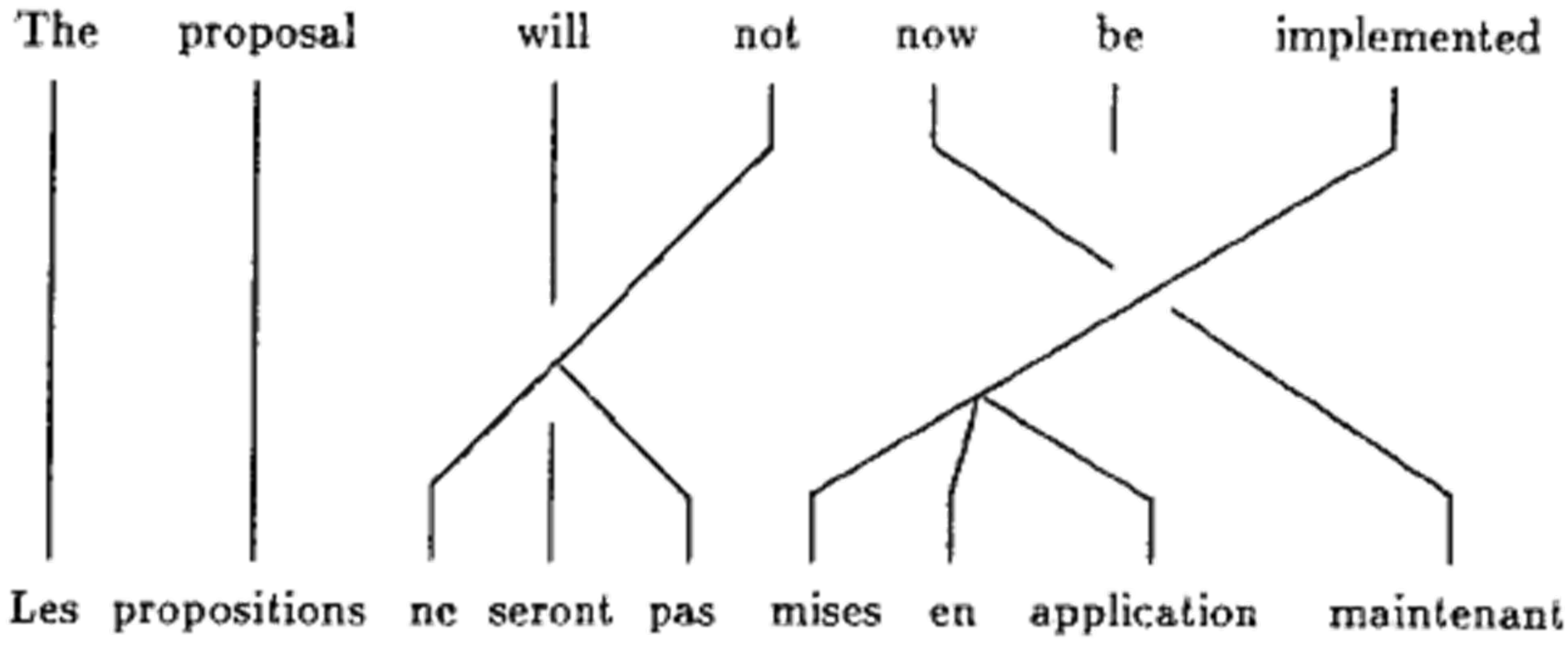
  In our organization research has two missions.

  In our missions research organization has two.

The translation model

- For simple sentences, it is reasonable to think of the French translation of an English sentence as being generated from the English sentence word by word.
- We refer to a picture such as that shown in Figure 3 as an alignment.
- An alignment indicates the origin in the English sentence of each of the words in the French sentence.

Figure 3


The
proposal
will
not
now
be
implemented
Les
propositions
ne
seront
pas
mises
en
application
maintenant

Fertility and distortion

- We call the number of French words that an English word produces in a given alignment its *fertility* in that alignment.
- Sometimes, a French word will appear quite far from the English word that produced it. We call this effect *distortion.*

Fertility probabilities

- $\Pr(n \mid e)$ for each English word e and for each fertility n from 0 to some moderate limit, in our case 25.

Fertility and translation

- (*Le chien est battu par Jean | John(6) does beat(3,4) the(1) dog(2)*)

$$\Pr(\mathit{fertility}=1\,|\,\mathit{John})\times\Pr(\mathit{Jean}\,|\,\mathit{John})\times$$
$$\Pr(\mathit{fertility}=0\,|\,\mathit{does})\times$$
$$\Pr(\mathit{fertility}=2\,|\,\mathit{beat})\times\Pr(\mathit{est}\,|\,\mathit{beat})\Pr(\mathit{battu}\,|\,\mathit{beat})\times$$
$$\Pr(\mathit{fertility}=1\,|\,\mathit{the})\times\Pr(\mathit{Le}\,|\,\mathit{the})\times$$
$$\Pr(\mathit{fertility}=1\,|\,\mathit{dog})\times\Pr(\mathit{chien}\,|\,\mathit{dog})\times$$
$$\Pr(\mathit{fertility}=1\,|\,\langle \mathit{null}\rangle)\times\Pr(\mathit{par}\,\langle \mathit{null}\rangle).$$

Distortion probabilities

- We assume that the position of the target word depends only on the length of the target sentence and the position of the source word.
- $\Pr(i \mid j, l)$ where i is a target position, j a source position, and l the target length.

Searching

- In searching for the sentence S that maximizes $\Pr(S)\Pr(T \mid S)$, we face the difficulty that there are simply too many sentences to try.
- Using a variant of the *stack search*
- (*Jean aime Marie* |*)

where the asterisk is a place holder for an unknown sequence of source words.

Searching

- The search proceeds by iterations, each of which extends some of the most promising entries on the list.
- An entry is extended by adding one or more additional words to its hypothesis.
- The search ends when there is a complete alignment on the list that is significantly more promising than any of the incomplete alignments.

Example

- We might extend the initial entry above to one or more of the following entries:
- (*Jean aime Marie* |*John(1*)*),

  (*Jean aime Marie* |**loves(2*)*),

  (*Jean aime Marie* |**Mary(3*)),

  (*Jean airne Marie* |*Jeans(1*)*).

Parameter estimation

- Hansards
- We could estimate the parameters of the translation model by counting.

Parameter estimation

- Given some initial estimate of the parameters, we can compute the probability of any particular alignment.
- We can then re-estimate the parameters by weighing each possible alignment according to its probability as determined by the initial guess of the parameters.
- Repeated iterations of this process lead to parameters that assign ever greater probability to the set of sentence pairs that we actually observe.

Two pilot experiments

- In our first experiment, we test our ability to estimate parameters for the translation model.
- Hansards
- 9,000 most common words
- 40,000 pairs of sentences
- Comprising 800,000 words

First experiment

- Figure 4 shows the translation and fertility probabilities we estimated for the English word *the.*

Figure 4

English: the

French	Probability	Fertility	Probability
le	.610	1	.871
la	.178	0	.124
l'	.083	2	.004
les	.023		
ce	.013		
il	.012		
de	.009		
à	.007		
que	.007		

First experiment

- In some sense, this correspondence is inherent in the sentence pairs themselves.
- Figure 5 shows these probabilities for the English word *not.*
- In Figure 6, we see the trained parameters for the English word *hear.*

Figure 5

English: not

French	**Probability**	**Fertility**	**Probability**
pas	.469	2	.758
ne	.460	0	.133
non	.024	1	.106
pas du tout	.003		
faux	.003		
plus	.002		
ce	.002		
que	.002		
jamais	.002		

Figure 6

English: hear

French	Probability	Fertility	Probability
bravo	.992	0	.584
entendre	.005	1	.416
entendu	.002		
entends	.001		

Second experiment

- We used the statistical approach to translate from French to English.
- Hansards
- English vocabulary: 1,000 words
- French vocabulary: 1,700 words

Model

- Translation model
  - Sentences: 117,000

- Bigram language model
  - Sentences: 570,000
  - Words: 12 million

Category

- Exact
- Alternate
- Different
- Wrong
- Ungrammatical

Translation Results

Category	Number of sentences	Percent
Exact	4	5
Alternate	18	25
Different	13	18
Wrong	11	15
Ungramatical	27	37
Total	73	

Translation Results

- As an alternate measure of the system's performance, one of us corrected each of the sentences in the last three categories (different, wrong, and ungrammatical) to either the exact or the alternate category.
- 776 strokes
- This compares with the 1,916 strokes required to generate all of the Hansard translations from scratch.
- the system reduces the work by about 60%.

Plans

- Expect
- Problems
- Trigram language model
- Morphologies