NLP

Machine Translation

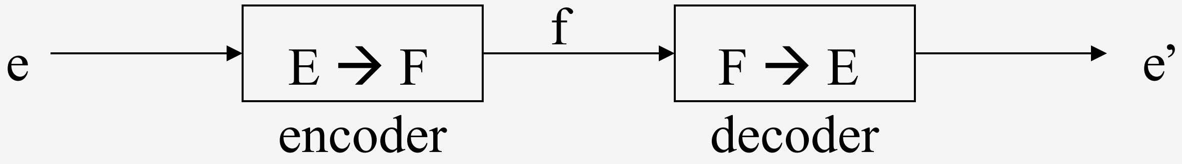
Noisy Channel Methods

The Noisy Channel Model

- Source-channel model of communication
- Parametric probabilistic models of language and translation

Statistics

- Given f, guess e



$$e' = \arg\max_{e} P(e|f) = \arg\max_{e} P(f|e)\, P(e)$$

translation model   language model

Statistical MT

Translate from French: “une fleur rouge”?

	p(e)	p(f|e)	p(e)*p(f|e)
a flower red			
red flower a			
flower red a			
a red dog			
dog cat mouse			
a red flower			

Statistical MT

	p(e)	p(f|e)	p(e)*p(f|e)
a flower red	low	high	low
red flower a			
flower red a			
a red dog			
dog cat mouse			
a red flower			

Statistical MT

	p(e)	p(f|e)	p(e)*p(f|e)
a flower red	low	high	low
red flower a	low	high	low
flower red a			
a red dog			
dog cat mouse			
a red flower			

Statistical MT

	p(e)	**p(f|e)**	**p(e)*p(f|e)**
a flower red	low	high	low
red flower a	low	high	low
flower red a	low	high	low
a red dog			
dog cat mouse			
a red flower			

Statistical MT

	p(e)	p(f|e)	p(e)*p(f|e)
a flower red	low	high	low
red flower a	low	high	low
flower red a	low	high	low
a red dog	high	low	low
dog cat mouse	low	low	low
a red flower	high	high	high

Example

$$p(\text{Chinese}|\text{English})$$
$$\times\ p(\text{English})$$
$$\sim p(\text{English}|\text{Chinese})$$

MT/Noisy Channel Models

- Text-to-text (summarization)
  - also text-to-signal, speech recognition, OCR, spelling correction
- Example (OCR)
  - $P(\text{text}|\text{pixels}) = P(\text{text})\,P(\text{pixels}|\text{text})$

Generative Story (almost IBM)

- I watched an interesting play
- I watched watched an interesting play play play
- I watched watched an play play play interesting
- J’ ai vu une pièce de théâtre intéressante

IBM's EM Trained Models (1-5)

- Word translation
- Local alignment
- Fertilities
- Class-based alignment
- Non-deficient algorithm (avoid overlaps, overflow)

Steps

- Tokenization
- Sentence alignment (1-1, 2-2, 2-1 mappings)
  - Church and Gale (based on sentence length)
  - Church (sequences of 4-grams) - based on cognates

Sentence Alignment

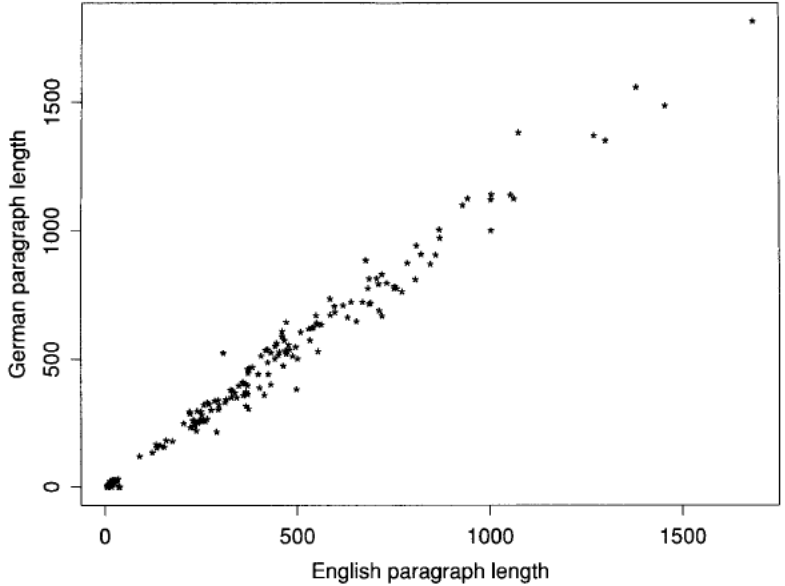


Figure 1
Paragraph lengths are highly correlated. The horizontal axis shows the length of English paragraphs, while the vertical scale shows the lengths of the corresponding German paragraphs. Note that the correlation is quite large (.991).

Table 5
Prob(match)

Category	Frequency	Prob(match)
1-1	1167	0.89
1-0 or 0-1	13	0.0099
2-1 or 1-2	117	0.089
2-2	15	0.011
	1312	1.00

[Church/Gale 1993]

Model 1

- Alignments
  - La maison bleue
  - The blue house
  - Alignments: {1,2,3}, {1,3,2}, {1,3,3}, {1,1,1}
  - All are equally likely
- Conditional probabilities
  - $P(f|A,e) = ?$

Model 1 (cont'd)

- Algorithm
  - Pick length of translation
  - Choose an alignment
  - Pick the French words
  - That gives you P(f,A|e)
  - We need P(f|A,e)
  - Use EM (expectation-maximisation) to find the hidden variables

Model 1

- We need p(f|e) but we don't know the word alignments (which are assumed to be equally likely)

$$p(f, A \mid e) = p(A \mid e) * p(f \mid A, e) = \frac{c}{(l+1)^m} \prod_{j=1}^{m} p(f_j \mid e_{a_j})$$

Model 2

- Distortion parameters D(i|j,l,m)
  - i and j are words in the two sentences
  - l and m are the lengths of these sentences.

Model 3

- Fertility
- $P(\phi_i|e)$
- Examples
  - (a) play = pièce de théâtre
  - (to) place = mettre en place
- p_1 is an extra parameter that defines ϕ_0

References

- http://www.isi.edu/natural-language/mt/wkbk.rtf

(an awesome tutorial by Kevin Knight)

- http://www.statmt.org/

(a comprehensive site, including references to the old IBM papers, pointers to Moses, for hw5, etc.

NLP