SFU NatLangLab

Natural Language Processing

Anoop Sarkar
anoopsarkar.github.io/nlp-class

Simon Fraser University

October 16, 2014

Natural Language Processing

Anoop Sarkar
anoopsarkar.github.io/nlp-class

Simon Fraser University

Part 1: Generative Models for Word Alignment

Statistical Machine Translation

Noisy Channel Model

$$\mathbf{e}^* = \arg\max_{\mathbf{e}} \underbrace{\Pr(\mathbf{e})}_{\textbf{Language Model}} \cdot \underbrace{\Pr(\mathbf{f} \mid \mathbf{e})}_{\textbf{Alignment Model}}$$

Alignment Task

$$\begin{array}{l} \mathbf{e} \searrow \\ \qquad \text{Program} \longrightarrow \Pr(\mathbf{e} \mid \mathbf{f}) \\ \mathbf{f} \nearrow \end{array}$$

Training Data

- **Alignment Model**: learn a mapping between **f** and **e**.
  Training data: lots of translation pairs between **f** and **e**.

Statistical Machine Translation

The IBM Models

- The first statistical machine translation models were developed at IBM Research (Yorktown Heights, NY) in the 1980s
- The models were published in 1993:
  Brown et. al. The Mathematics of Statistical Machine Translation.
  Computational Linguistics. 1993.
  `http://aclweb.org/anthology/J/J93/J93-2003.pdf`
- These models are the basic SMT models, called:
  IBM Model 1, IBM Model 2, IBM Model 3, IBM Model 4, IBM Model 5
  as they were called in the 1993 paper.
- We use **e** and **f** in the equations in honor of their system which translated from French to English.
  Trained on the Canadian Hansards (Parliament Proceedings)

Statistical Machine Translation

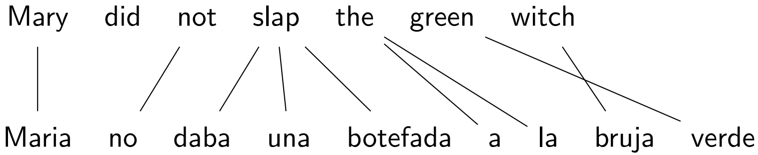
Generative Model of Word Alignment

- Word Alignments: IBM Model 3
- Word Alignments: IBM Model 1

Generative Model of Word Alignment

- English **e**: Mary did not slap the green witch
- "French" **f**: Maria no daba una botefada a la bruja verde
- Alignment **a**: $\{1, 3, 4, 4, 4, 5, 5, 7, 6\}$
  e.g. $(f_8, e_{a_8}) = (f_8, e_7) =$ (bruja, witch)

Visualizing alignment **a**



Generative Model of Word Alignment

Data Set

- Data set $\mathcal{D}$ of N sentences:
$$\mathcal{D} = \{(\mathbf{f}^{(1)}, \mathbf{e}^{(1)}), \ldots, (\mathbf{f}^{(N)}, \mathbf{e}^{(N)})\}$$
- French $\mathbf{f}$: $(f_1, f_2, \ldots, f_I)$
- English $\mathbf{e}$: $(e_1, e_2, \ldots, e_J)$
- Alignment $\mathbf{a}$: $(a_1, a_2, \ldots, a_I)$

Generative Model of Word Alignment

Find the best alignment for each translation pair

$$\mathbf{a}^* = \arg\max_{\mathbf{a}} \Pr(\mathbf{a} \mid \mathbf{f}, \mathbf{e})$$

Chain rule revisited

$$\begin{aligned} \Pr(\mathbf{f}, \mathbf{a}) &= \Pr(\mathbf{f}) \Pr(\mathbf{a} \mid \mathbf{f}) \\ \Pr(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) &= \Pr(\mathbf{f} \mid \mathbf{e}) \Pr(\mathbf{a} \mid \mathbf{f}, \mathbf{e}) \end{aligned}$$

Alignment probability

$$\Pr(\mathbf{a} \mid \mathbf{f}, \mathbf{e}) = \frac{\Pr(\mathbf{f}, \mathbf{a} \mid \mathbf{e})}{\Pr(\mathbf{f} \mid \mathbf{e})} = \frac{\Pr(\mathbf{f}, \mathbf{a} \mid \mathbf{e})}{\sum_{\mathbf{a}} \Pr(\mathbf{f}, \mathbf{a} \mid \mathbf{e})}$$

Statistical Machine Translation

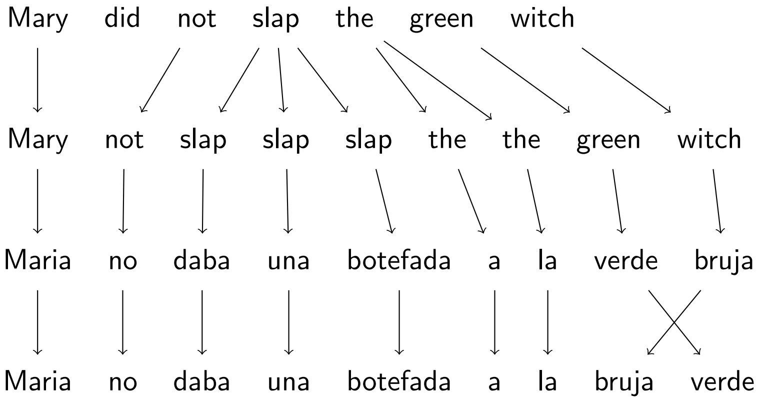
Generative Model of Word Alignment

Word Alignments: IBM Model 3

Word Alignments: IBM Model 1

Word Alignments: IBM Model 3

Generative "story" for $P(\mathbf{a} \mid \mathbf{f}, \mathbf{e})$



Word Alignments: IBM Model 3

Fertility parameter

$$n(\phi_j \mid e_j) : n(3 \mid slap)$$

Translation parameter

$$t(f_i \mid e_j) : t(bruja \mid witch)$$

Distortion parameter

$$d(f_{pos} \mid e_{pos}, I, J) : d(8 \mid 7, 8, 6)$$

Word Alignments: IBM Model 1

Alignment probability

$$\Pr(\mathbf{a} \mid \mathbf{f}, \mathbf{e}) = \frac{\Pr(\mathbf{f}, \mathbf{a} \mid \mathbf{e})}{\sum_{\mathbf{a}} \Pr(\mathbf{f}, \mathbf{a} \mid \mathbf{e})}$$

Example alignment

1	2	3	4
das	Haus	ist	klein
|	|	|	|
1	2	3	4
the	house	is	small

$$\Pr(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) = \prod_{i=1}^{I} t(f_i \mid e_{a_i})$$

$$\begin{aligned}\Pr(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) = \\ & t(\text{das} \mid \text{the}) \times \\ & t(\text{Haus} \mid \text{house}) \times \\ & t(\text{ist} \mid \text{is}) \times \\ & t(\text{klein} \mid \text{small})\end{aligned}$$

Word Alignments: IBM Model 1

Sum over all alignments

$$\sum_{\mathbf{a}} \Pr(\mathbf{f}, \mathbf{a} \mid \mathbf{e}) \quad = \quad \sum_{a_1=1}^{J} \sum_{a_2=1}^{J} \ldots \sum_{a_I=1}^{J} \prod_{i=1}^{I} t(f_i \mid e_{a_i})$$

Assume (f_1, f_2, f_3) and (e_1, e_2)

$$\sum_{a_1=1}^{2} \sum_{a_2=1}^{2} \sum_{a_3=1}^{2} t(f_1 \mid e_{a_1}) \times t(f_2 \mid e_{a_2}) \times t(f_3 \mid e_{a_3})$$

Word Alignments: IBM Model 1

Assume (f_1, f_2, f_3) and (e_1, e_2): $I = 3$ and $J = 2$

$$\sum_{a_1=1}^{2} \sum_{a_2=1}^{2} \sum_{a_3=1}^{2} t(f_1 \mid e_{a_1}) \times t(f_2 \mid e_{a_2}) \times t(f_3 \mid e_{a_3})$$

$J^I = 2^3$ terms to be added:

$t(f_1 \mid e_1)$	$\times$	$t(f_2 \mid e_1)$	$\times$	$t(f_3 \mid e_1)$	$+$
$t(f_1 \mid e_1)$	$\times$	$t(f_2 \mid e_1)$	$\times$	$t(f_3 \mid e_2)$	$+$
$t(f_1 \mid e_1)$	$\times$	$t(f_2 \mid e_2)$	$\times$	$t(f_3 \mid e_1)$	$+$
$t(f_1 \mid e_1)$	$\times$	$t(f_2 \mid e_2)$	$\times$	$t(f_3 \mid e_2)$	$+$
$t(f_1 \mid e_2)$	$\times$	$t(f_2 \mid e_1)$	$\times$	$t(f_3 \mid e_1)$	$+$
$t(f_1 \mid e_2)$	$\times$	$t(f_2 \mid e_1)$	$\times$	$t(f_3 \mid e_2)$	$+$
$t(f_1 \mid e_2)$	$\times$	$t(f_2 \mid e_2)$	$\times$	$t(f_3 \mid e_1)$	$+$
$t(f_1 \mid e_2)$	$\times$	$t(f_2 \mid e_2)$	$\times$	$t(f_3 \mid e_2)$	$+$

Word Alignments: IBM Model 1

Factor the terms:

$$\begin{array}{lcll}
(t(f_1 \mid e_1) \times t(f_2 \mid e_1)) & \times & (t(f_3 \mid e_1) + t(f_3 \mid e_2)) & + \\
(t(f_1 \mid e_1) \times t(f_2 \mid e_2)) & \times & (t(f_3 \mid e_1) + t(f_3 \mid e_2)) & + \\
(t(f_1 \mid e_2) \times t(f_2 \mid e_1)) & \times & (t(f_3 \mid e_1) + t(f_3 \mid e_2)) & + \\
(t(f_1 \mid e_2) \times t(f_2 \mid e_2)) & \times & (t(f_3 \mid e_1) + t(f_3 \mid e_2)) &
\end{array}$$

$$(t(f_3 \mid e_1) + t(f_3 \mid e_2)) \left(\begin{array}{lcll}
t(f_1 \mid e_1) & \times & t(f_2 \mid e_1) & + \\
t(f_1 \mid e_1) & \times & t(f_2 \mid e_2) & + \\
t(f_1 \mid e_2) & \times & t(f_2 \mid e_1) & + \\
t(f_1 \mid e_2) & \times & t(f_2 \mid e_2) &
\end{array} \right)$$

$$(t(f_3 \mid e_1) + t(f_3 \mid e_2)) \left(\begin{array}{lcll}
t(f_1 \mid e_1) & \times & (t(f_2 \mid e_1) + t(f_2 \mid e_2)) & + \\
t(f_1 \mid e_2) & \times & (t(f_2 \mid e_1) + t(f_2 \mid e_2)) &
\end{array} \right)$$

Word Alignments: IBM Model 1

Assume (f_1, f_2, f_3) and (e_1, e_2): $I = 3$ and $J = 2$

$$\prod_{i=1}^{3} \sum_{a_i=1}^{2} t(f_i \mid e_{a_i})$$

$I \times J = 2 \times 3$ terms to be added:

$$\begin{array}{ccccc}
(t(f_1 \mid e_1) & + & t(f_1 \mid e_2)) & \times \\
(t(f_2 \mid e_1) & + & t(f_2 \mid e_2)) & \times \\
(t(f_3 \mid e_1) & + & t(f_3 \mid e_2)) &
\end{array}$$

Acknowledgements

Many slides borrowed or inspired from lecture notes by Michael Collins, Chris Dyer, Kevin Knight, Adam Lopez, and Luke Zettlemoyer from their NLP course materials. All mistakes are my own.