Statistical Machine Translation with Weighted Grammars

Matthias Büchse

Chair of Foundations of Programming
Institute of Theoretical Computer Science
Technische Universität Dresden

June 15, 2011

Outline

Ambiguity in Natural Language

Example (Chiang 2007)

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

Ambiguity in Natural Language

Example (Chiang 2007)

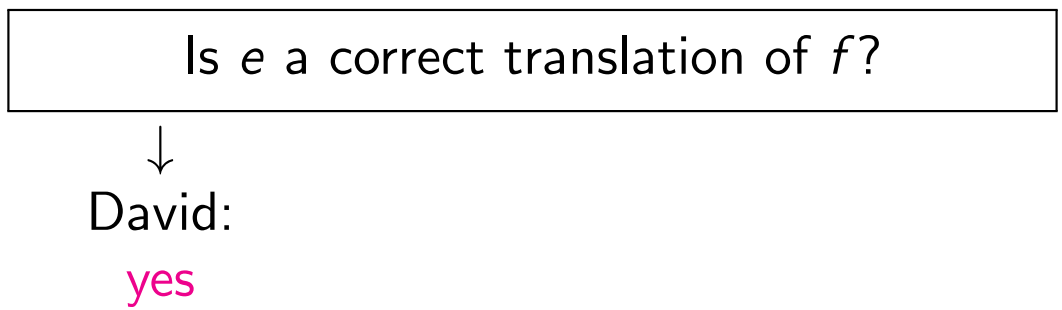
- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

Is e a correct translation of f?

Ambiguity in Natural Language

Example (Chiang 2007)

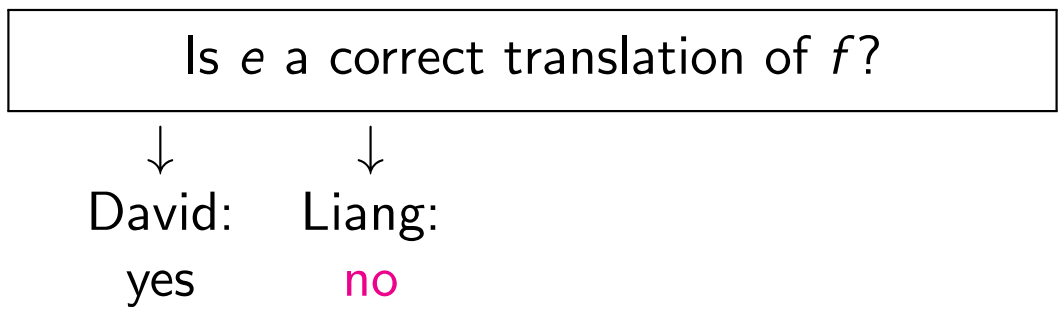
- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"



Ambiguity in Natural Language

Example (Chiang 2007)

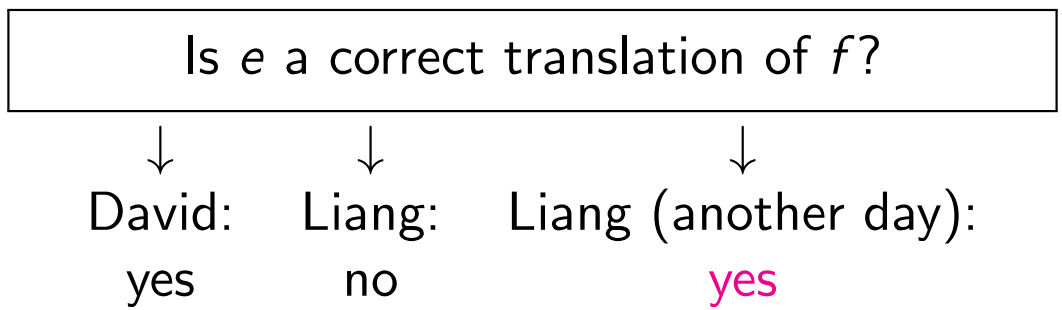
- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"



Ambiguity in Natural Language

Example (Chiang 2007)

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"



Ambiguity in Natural Language

Example (Chiang 2007)

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

What is $P(e \mid f)$?

Ambiguity in Natural Language

Example (Chiang 2007)

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

What is $P(e \mid f)$?

↓

.42

Ambiguity in Natural Language

Example (Chiang 2007)

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

What is $P(e \mid f)$?

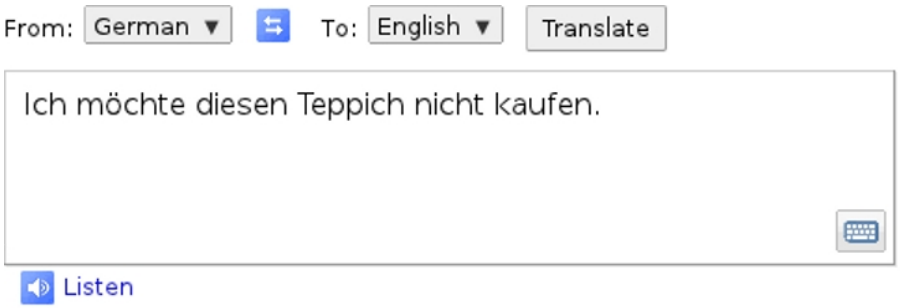
↓

.42

frequentist limit of relative frequencies

Bayesian degree of belief (subjective)

Search (aka Decoding) in SMT



given $p \in \mathcal{M}(E \mid F)$

Search (aka Decoding) in SMT

given $p \in \mathcal{M}(E \mid F)$ $(\sum_e p(e \mid f) = 1$ for every $f)$

Search (aka Decoding) in SMT

given $p \in \mathcal{M}(E \mid F)$ and $f \in F$,

Search (aka Decoding) in SMT

given $p \in \mathcal{M}(E \mid F)$ and $f \in F$,

determine $\hat{e}$ such that for every $e \in E$:

$$p(\hat{e} \mid f) \geq p(e \mid f)\ . \qquad \text{(not unique)}$$

Search (aka Decoding) in SMT

given $p \in \mathcal{M}(E \mid F)$ and $f \in F$,

determine $\hat{e}$ such that for every $e \in E$:

$$p(\hat{e} \mid f) \geq p(e \mid f) \; . \qquad \text{(not unique)}$$

for short

$$\hat{e} = \operatorname{argmax}_{e \in E} p(e \mid f) \; .$$

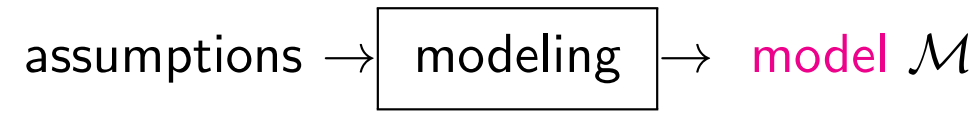
Building an SMT System

assumptions

training data

test data

Building an SMT System

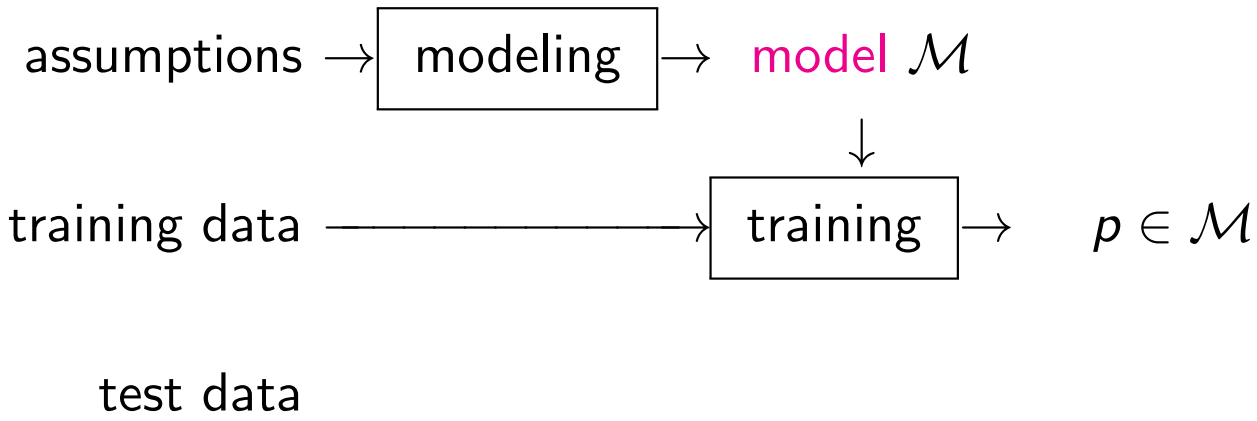


training data

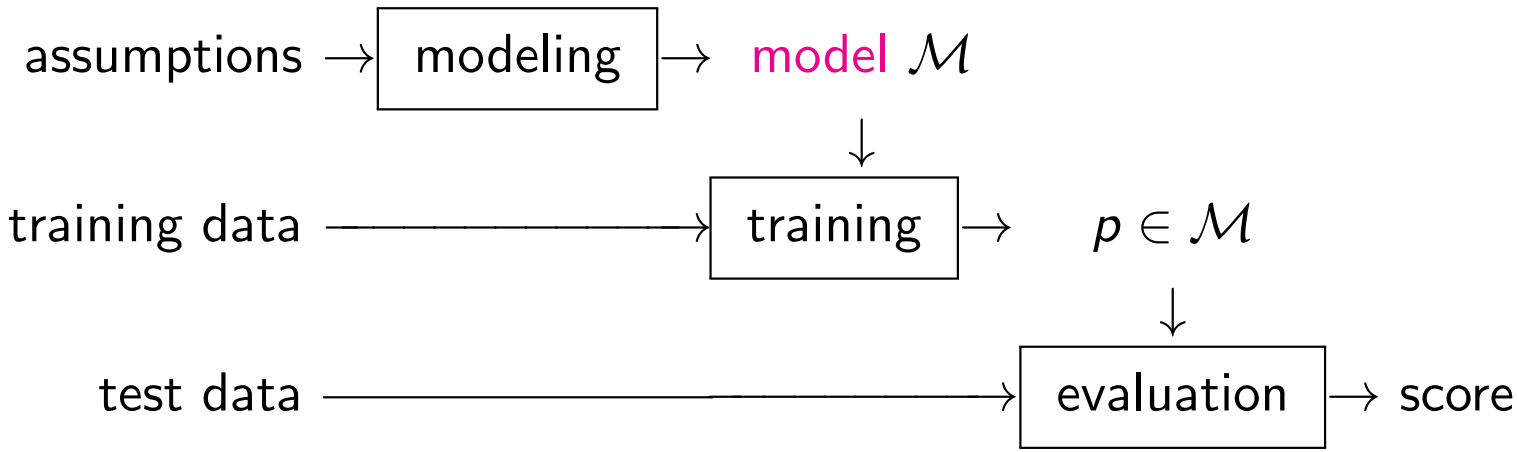
test data

"Every time I fire a linguist, recognition rate goes up"
(attributed to Frederick Jelinek)

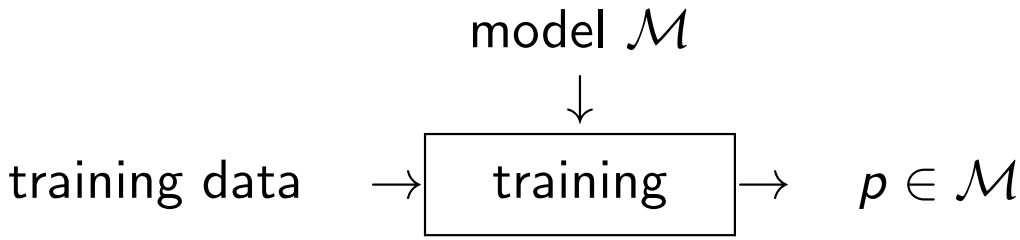
Building an SMT System



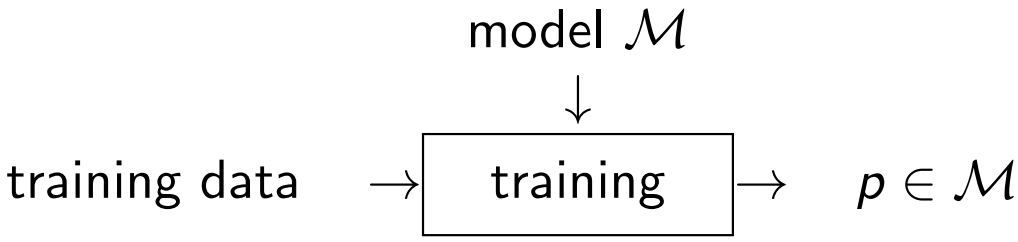
Building an SMT System

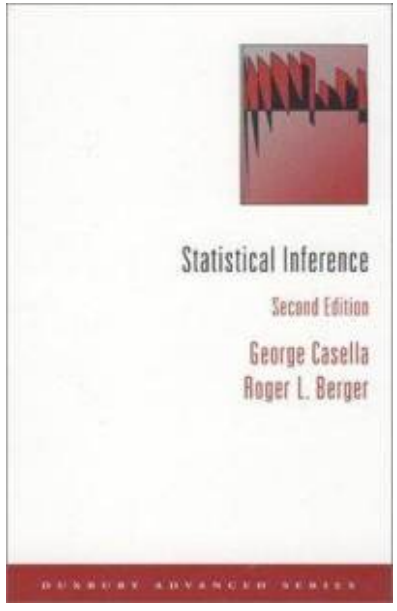


Training = Point Estimation



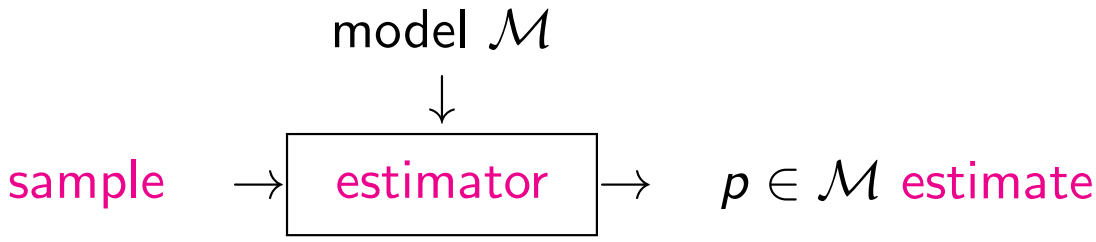
Training = Point Estimation

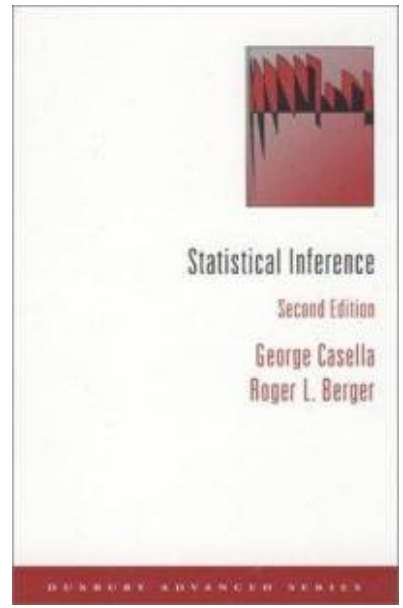




Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

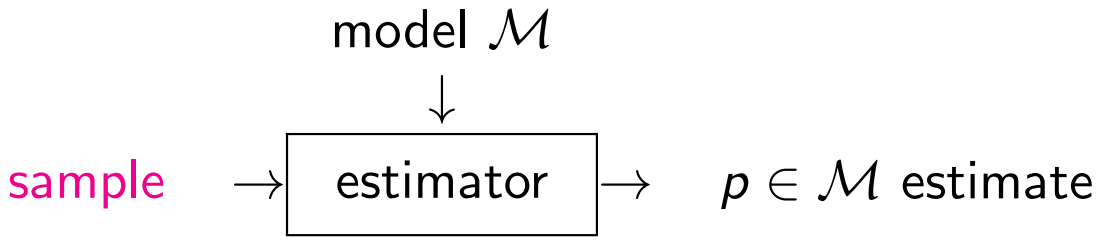
Training = Point Estimation





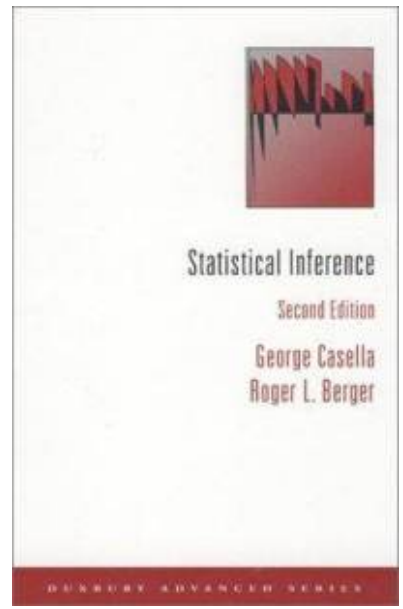
Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

Training = Point Estimation



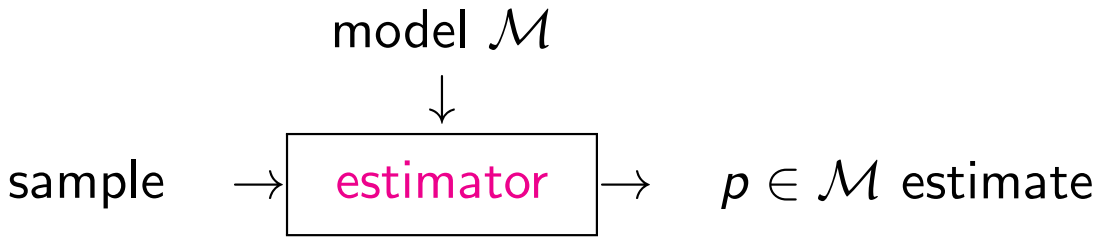
sample

- representative of all translation situations
- e. g., Hong Kong Hansards (parliament proceedings)



Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

Training = Point Estimation

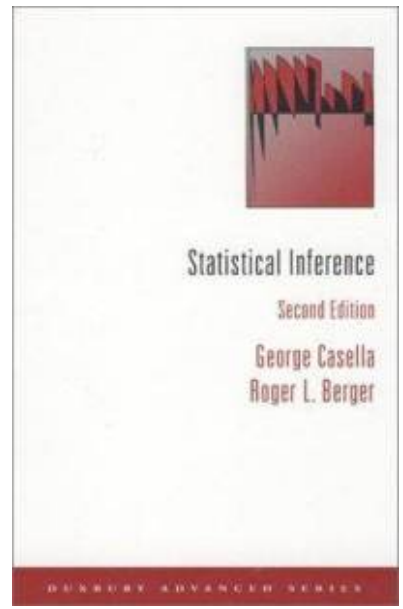


sample

- representative of all translation situations
- e. g., Hong Kong Hansards (parliament proceedings)

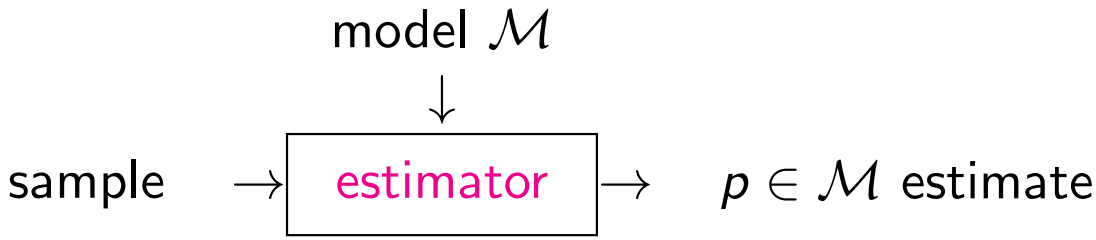
estimators

- maximum likelihood
- maximum a-posteriori
- minimum risk
- maximum entropy



Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

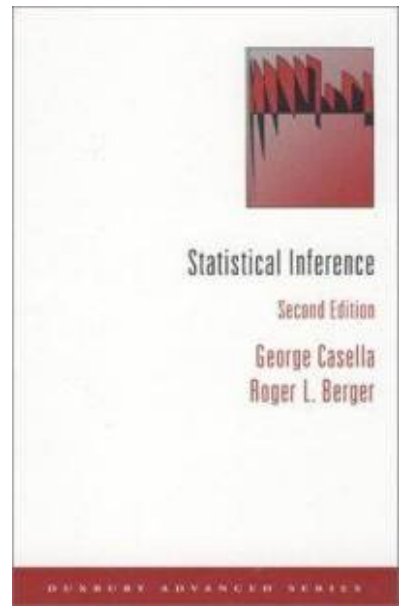
Training = Point Estimation



sample
- representative of all translation situations
- e. g., Hong Kong Hansards (parliament proceedings)

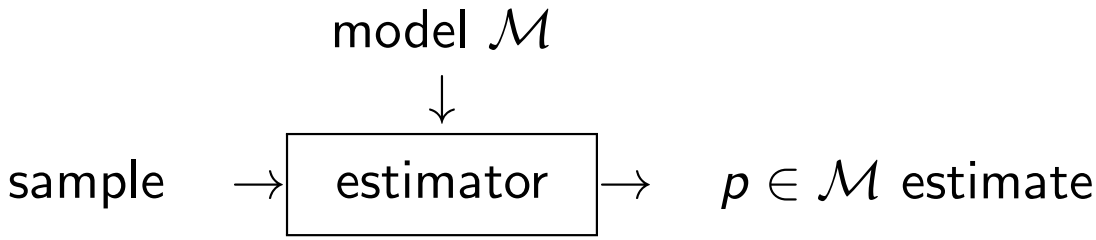
estimators
- maximum likelihood
- maximum a-posteriori
- minimum risk
- maximum entropy

properties
- loss function optimality (Statistical Decision Theory)
- asymptotic optimality (consistency)



Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

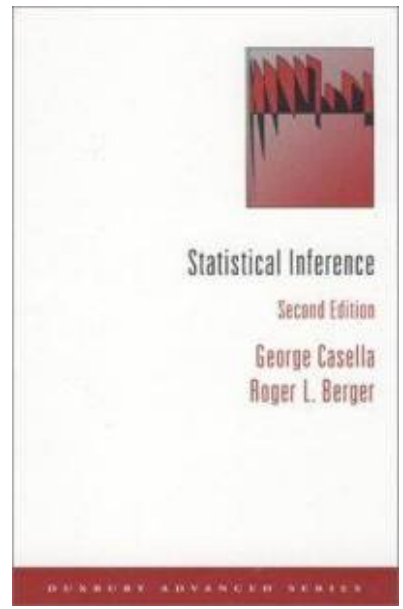
Training = Point Estimation



sample
- representative of all translation situations
- e. g., Hong Kong Hansards (parliament proceedings)

estimators
- maximum likelihood
- maximum a-posteriori
- minimum risk
- maximum entropy

properties
- loss function optimality (Statistical Decision Theory)
- asymptotic optimality (consistency)



Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

Sample?

- data sparsity

Sample?

- data sparsity
- noise

Sample?

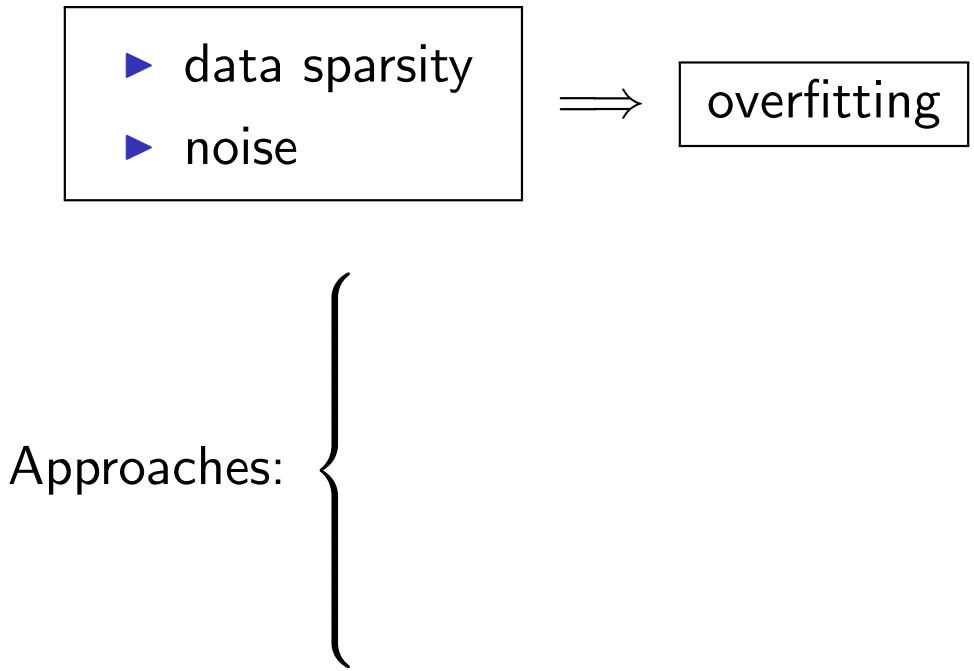
- data sparsity
- noise

Sample?

- data sparsity
- noise

$\Longrightarrow$ overfitting

Sample?



Sample?

- data sparsity
- noise

$\Longrightarrow$ overfitting

Approaches:

- cross validation

Sample?

- data sparsity
- noise

$\Longrightarrow$ overfitting

Approaches:

- cross validation
- use prior (smoothing)

Sample?

- data sparsity
- noise

$\Longrightarrow$ overfitting

Approaches:

- cross validation
- use prior (smoothing)
- use ad-hoc estimator (heuristics)

Sample?

- data sparsity
- noise

$\Longrightarrow$ overfitting

Approaches:

- cross validation
- use prior (smoothing)
- use ad-hoc estimator (heuristics)
- adjust the model

Adjusting the Model

Source-Channel Model (Brown et al. 1993)

- $\mathcal{M}_{\mathsf{LM}} \subseteq \mathcal{M}(E)$ language model
- $\mathcal{M}_{\mathsf{TM}} \subseteq \mathcal{M}(F \mid E)$ translation model

Adjusting the Model

Source-Channel Model (Brown et al. 1993)

- $\mathcal{M}_{\text{LM}} \subseteq \mathcal{M}(E)$ language model
- $\mathcal{M}_{\text{TM}} \subseteq \mathcal{M}(F \mid E)$ translation model
- then

$$\mathcal{M} = \{p \in \mathcal{M}(E \mid F) \;\mid\; \exists p_1 \in \mathcal{M}_{\text{LM}}, p_2 \in \mathcal{M}_{\text{TM}}: \; p(e \mid f) = \frac{p_1(e) \cdot p_2(f \mid e)}{\sum_{e' \in E} p_1(e') \cdot p_2(f \mid e')}\}$$

Adjusting the Model

Source-Channel Model (Brown et al. 1993)

- $\mathcal{M}_{\text{LM}} \subseteq \mathcal{M}(E)$ language model
- $\mathcal{M}_{\text{TM}} \subseteq \mathcal{M}(F \mid E)$ translation model
- then

$$\mathcal{M} = \{p \in \mathcal{M}(E \mid F) \;\; \Big| \;\; \exists p_1 \in \mathcal{M}_{\text{LM}}, p_2 \in \mathcal{M}_{\text{TM}}: \\ p(e \mid f) = \frac{p_1(e) \cdot p_2(f \mid e)}{\sum_{e' \in E} p_1(e') \cdot p_2(f \mid e')}\}$$

- estimate p_1 and p_2 independently on appropriate data

Adjusting the Model

Log-linear Model (Berger, Della Pietra, and Della Pietra 1996; Och and Ney 2002)

- $h_1, \ldots, h_n \colon E \times F \to \mathbb{R}$ features

Adjusting the Model

Log-linear Model (Berger, Della Pietra, and Della Pietra 1996; Och and Ney 2002)

- $h_1, \ldots, h_n \colon E \times F \to \mathbb{R}$ features
- consider

$$\mathcal{M} = \{p \in \mathcal{M}(E \mid F) \;\; \Big| \;\; \forall i \colon \mathbb{E}_p[h_i] = \mathbb{E}_{\tilde{p}}[h_i]\}$$

- (and maximize entropy)

Adjusting the Model

Log-linear Model (Berger, Della Pietra, and Della Pietra 1996; Och and Ney 2002)

- $h_1, \ldots, h_n \colon E \times F \to \mathbb{R}$ features
- consider

$$\mathcal{M} = \{p \in \mathcal{M}(E \mid F) \;\mid\; \forall i \colon \mathbb{E}_p[h_i] = \mathbb{E}_{\tilde{p}}[h_i]\}$$

- (and maximize entropy)
- or consider

$$\mathcal{M} = \{p \in \mathcal{M}(E \mid F) \;\mid\; \exists \lambda_1, \ldots, \lambda_n \in \mathbb{R} \colon p(e \mid f) = \frac{\exp \sum_i \lambda_i \cdot h_i(e, f)}{\sum_{e' \in E} \exp \sum_i \lambda_i \cdot h_i(e', f)}\}$$

- (and maximize likelihood)

Adjusting the Model

Log-linear Model Example

- $h_1(e, f) = \log p_1(e)$,
- $h_2(e, f) = \log p_2(f \mid e)$,
- $\lambda_1 = 1$,
- $\lambda_2 = 1$,
- then

$$
\begin{aligned}
p(e \mid f) &= \frac{\exp\big(\lambda_1 \cdot h_1(e, f) + \lambda_2 \cdot h_2(e, f)\big)}{\sum_{e' \in E} \exp\big(\lambda_1 \cdot h_1(e', f) + \lambda_2 \cdot h_2(e, f)\big)} \\
&= \frac{\exp\big(\log p_1(e) + \log p_2(f \mid e)\big)}{\sum_{e' \in E} \exp\big(\log p_1(e') + \log p_2(f \mid e)\big)} \\
&= \frac{p_1(e) \cdot p_2(f \mid e)}{\sum_{e' \in E} p_1(e') \cdot p_2(f \mid e)}
\end{aligned}
$$

Estimating Conditional Models

- $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- define

$$\mathcal{M}' = \{p' \in \mathcal{M}(E, F) \mid \exists p \in \mathcal{M}, p_F \in \mathcal{M}(F): \\ p'(e, f) = p(e \mid f) \cdot p_F(f)\}$$

Estimating Conditional Models

- $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- define

$$\mathcal{M}' = \{p' \in \mathcal{M}(E, F) \mid \exists p \in \mathcal{M}, p_F \in \mathcal{M}(F): \\ p'(e, f) = p(e \mid f) \cdot p_F(f)\}$$

- bijection: $p' \leftrightarrow (p, p_F)$ iff $p'(e, f) = p(e \mid f) \cdot p_F(f)$
- if $P(e, f) = p'(e, f)$, then $P(e \mid f) = p(e \mid f)$,
- but p and p_F are completely independent,

Estimating Conditional Models

- $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- define

$$\mathcal{M}' = \{p' \in \mathcal{M}(E, F) \mid \exists p \in \mathcal{M}, p_F \in \mathcal{M}(F): \\ p'(e, f) = p(e \mid f) \cdot p_F(f)\}$$

- bijection: $p' \leftrightarrow (p, p_F)$ iff $p'(e, f) = p(e \mid f) \cdot p_F(f)$
- if $P(e, f) = p'(e, f)$, then $P(e \mid f) = p(e \mid f)$,
- but p and p_F are completely independent,

$\rightsquigarrow$ do estimation on $\mathcal{M}'$, throw p_F away

More Recent Estimators

- minimum-error-rate training (Och 2003)
- minimum-risk annealing (Smith and Eisner 2006)
- large-margin methods

Evaluation

(due to Saša Hasan)

Evaluation

$$p \in \mathcal{M}$$

$$\text{test data} \rightarrow \boxed{\text{evaluation}} \rightarrow \text{score}$$

Evaluation

$$p \in \mathcal{M}$$

$$\downarrow$$

$$\text{test data} \rightarrow \boxed{\text{evaluation}} \rightarrow \text{score}$$

$$\text{training data} \cap \text{test data} = \emptyset$$

Evaluation

$$p \in \mathcal{M}$$

$$\downarrow$$

$$\text{test data} \rightarrow \boxed{\text{evaluation}} \rightarrow \text{score}$$

$$\text{training data} \cap \text{test data} = \emptyset$$

method	how it works	goal
BLEU (Papineni et al. 2002)	n-gram precision, brevity penalty	higher values

Evaluation

$$p \in \mathcal{M}$$

$$\downarrow$$

$$\text{test data} \rightarrow \boxed{\text{evaluation}} \rightarrow \text{score}$$

$$\text{training data} \cap \text{test data} = \emptyset$$

method	how it works	goal
BLEU (Papineni et al. 2002)	n-gram precision, brevity penalty	higher values
TER (Snover et al. 2006)	edit distance	lower values

Evaluation

$$p \in \mathcal{M}$$

$$\text{test data} \rightarrow \boxed{\text{evaluation}} \rightarrow \text{score}$$

$$\text{training data} \cap \text{test data} = \emptyset$$

method	how it works	goal
BLEU (Papineni et al. 2002)	n-gram precision, brevity penalty	higher values
TER (Snover et al. 2006)	edit distance	lower values
Meteor (Banerjee and Lavie 2005)	alignments, unigram precision, unigram recall, penalty	higher values

Outline

Example (Chiang 2007)

Recall

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

Example (Chiang 2007)

Recall

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

Synchronous Context-Free Grammar G:

π_1: $\mathrm{S} \rightarrow \langle \mathrm{S}\ \mathrm{X},\ \mathrm{S}\ \mathrm{X} \rangle$

π_2: $\mathrm{S} \rightarrow \langle \mathrm{X},\ \mathrm{X} \rangle$

π_3: $\mathrm{X} \rightarrow \langle \text{yu}\ \mathrm{X}_{\boxed{1}}\ \text{you}\ \mathrm{X}_{\boxed{2}},\ \text{have}\ \mathrm{X}_{\boxed{2}}\ \text{with}\ \mathrm{X}_{\boxed{1}} \rangle$

π_4: $\mathrm{X} \rightarrow \langle \mathrm{X}_{\boxed{1}}\ \text{de}\ \mathrm{X}_{\boxed{2}},\ \text{the}\ \mathrm{X}_{\boxed{2}}\ \text{that}\ \mathrm{X}_{\boxed{1}} \rangle$

π_5: $\mathrm{X} \rightarrow \langle \mathrm{X}\ \text{zhiyi},\ \text{one of}\ \mathrm{X} \rangle$

π_6: $\mathrm{X} \rightarrow \langle \text{Aozhou},\ \text{Australia} \rangle$

π_7: $\mathrm{X} \rightarrow \langle \text{Beihan},\ \text{North Korea} \rangle$

π_8: $\mathrm{X} \rightarrow \langle \text{shi},\ \text{is} \rangle$

π_9: $\mathrm{X} \rightarrow \langle \text{bangjiao},\ \text{diplomatic relations} \rangle$

π_{10}: $\mathrm{X} \rightarrow \langle \text{shaoshu guojia},\ \text{few countries} \rangle$

Derivation

$\langle S, S \rangle$

Derivation

$$\langle \mathsf{S},\ \mathsf{S} \rangle$$
$$\overset{\pi_1}{\Rightarrow}\ \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle$$

Derivation

$$\begin{aligned} & \langle \mathsf{S},\ \mathsf{S} \rangle \\ \overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \end{aligned}$$

Derivation

$$
\begin{aligned}
& \langle \mathsf{S},\ \mathsf{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathsf{S},\ \mathsf{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathsf{S},\ \mathsf{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathsf{S},\ \mathsf{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathsf{S},\ \mathsf{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \text{Aozhou}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \text{Australia}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X},\ \mathrm{S}\ \mathrm{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X},\ \mathrm{S}\ \mathrm{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \text{Aozhou}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \text{Australia}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \text{Aozhou shi X},\ \text{Australia is X} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S}, \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow} \; & \langle \mathrm{S}\ \mathrm{X}, \mathrm{S}\ \mathrm{X} \rangle \\
\overset{\pi_1}{\Rightarrow} \; & \langle \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow} \; & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow} \; & \langle \text{Aozhou}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \text{Australia}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow} \; & \langle \text{Aozhou shi X}, \text{Australia is X} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S}, \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X}, \mathrm{S}\ \mathrm{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \text{Aozhou}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \text{Australia}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \text{Aozhou shi X}, \text{Australia is X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \text{Aozhou shi X zhiyi}, \text{Australia is one of X} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X}_{\boxed{1}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \ \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ X}_{\boxed{1}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X}_{\boxed{1}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \ \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ X}_{\boxed{1}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X}_{\boxed{1}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ X}_{\boxed{1}} \rangle \\
\overset{\pi_3}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ X}_{\boxed{1}}\ \mathrm{you\ X}_{\boxed{0}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ have\ X}_{\boxed{0}}\ \mathrm{with\ X}_{\boxed{1}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X}_{\boxed{1}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ X}_{\boxed{1}} \rangle \\
\overset{\pi_3}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ X}_{\boxed{1}}\ \mathrm{you\ X}_{\boxed{0}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ have\ X}_{\boxed{0}}\ \mathrm{with\ X}_{\boxed{1}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X}_{\boxed{1}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ X}_{\boxed{1}} \rangle \\
\overset{\pi_3}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ X}_{\boxed{1}}\ \mathrm{you\ X}_{\boxed{0}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ have\ X}_{\boxed{0}}\ \mathrm{with\ X}_{\boxed{1}} \rangle \\
\overset{\pi_7}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ Beihan\ you\ X}_{\boxed{0}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ have\ X}_{\boxed{0}}\ \mathrm{with\ North\ Korea} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{S\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{Australia\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X_{\boxed{1}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ X_{\boxed{1}}} \rangle \\
\overset{\pi_3}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ X_{\boxed{1}}\ you\ X_{\boxed{0}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ have\ X_{\boxed{0}}\ with\ X_{\boxed{1}}} \rangle \\
\overset{\pi_7}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ Beihan\ you\ X_{\boxed{0}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ have\ X_{\boxed{0}}\ with\ North\ Korea} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{S\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{Australia\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X_{\boxed{1}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ X_{\boxed{1}}} \rangle \\
\overset{\pi_3}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ X_{\boxed{1}}\ you\ X_{\boxed{0}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ have\ X_{\boxed{0}}\ with\ X_{\boxed{1}}} \rangle \\
\overset{\pi_7}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ Beihan\ you\ X_{\boxed{0}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ have\ X_{\boxed{0}}\ with\ North\ Korea} \rangle \\
\overset{\pi_9}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ Beihan\ you\ bangjiao\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ have\ diplomatic\ relations\ with\ North\ Korea} \rangle
\end{aligned}
$$

Derivation

$\langle \mathrm{S}, \mathrm{S} \rangle$

$\overset{\pi_1}{\Rightarrow} \langle \mathrm{S}\ \mathrm{X}, \mathrm{S}\ \mathrm{X} \rangle$

$\overset{\pi_1}{\Rightarrow} \langle \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle$

$\overset{\pi_2}{\Rightarrow} \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle$

$\overset{\pi_6}{\Rightarrow} \langle \text{Aozhou } \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \text{Australia } \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle$

$\overset{\pi_8}{\Rightarrow} \langle \text{Aozhou shi X, Australia is X} \rangle$

$\overset{\pi_5}{\Rightarrow} \langle \text{Aozhou shi X zhiyi, Australia is one of X} \rangle$

$\overset{\pi_4}{\Rightarrow} \langle \text{Aozhou shi } \mathrm{X}_{\boxed{1}} \text{ de } \mathrm{X}_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } \mathrm{X}_{\boxed{2}} \text{ that } \mathrm{X}_{\boxed{1}} \rangle$

$\overset{\pi_3}{\Rightarrow} \langle \text{Aozhou shi yu } \mathrm{X}_{\boxed{1}} \text{ you } \mathrm{X}_{\boxed{0}} \text{ de } \mathrm{X}_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } \mathrm{X}_{\boxed{2}} \text{ that have } \mathrm{X}_{\boxed{0}} \text{ with } \mathrm{X}_{\boxed{1}} \rangle$

$\overset{\pi_7}{\Rightarrow} \langle \text{Aozhou shi yu Beihan you } \mathrm{X}_{\boxed{0}} \text{ de } \mathrm{X}_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } \mathrm{X}_{\boxed{2}} \text{ that have } \mathrm{X}_{\boxed{0}} \text{ with North Korea} \rangle$

$\overset{\pi_9}{\Rightarrow} \langle \text{Aozhou shi yu Beihan you bangjiao de } \mathrm{X}_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } \mathrm{X}_{\boxed{2}} \text{ that have diplomatic relations with North Korea} \rangle$

Derivation

$\langle S, S \rangle$

$\overset{\pi_1}{\Rightarrow} \langle S\ X, S\ X \rangle$

$\overset{\pi_1}{\Rightarrow} \langle S\ X_{\boxed{1}}\ X_{\boxed{2}}, S\ X_{\boxed{1}}\ X_{\boxed{2}} \rangle$

$\overset{\pi_2}{\Rightarrow} \langle X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}, X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}} \rangle$

$\overset{\pi_6}{\Rightarrow} \langle \text{Aozhou } X_{\boxed{1}}\ X_{\boxed{2}}, \text{Australia } X_{\boxed{1}}\ X_{\boxed{2}} \rangle$

$\overset{\pi_8}{\Rightarrow} \langle \text{Aozhou shi X, Australia is X} \rangle$

$\overset{\pi_5}{\Rightarrow} \langle \text{Aozhou shi X zhiyi, Australia is one of X} \rangle$

$\overset{\pi_4}{\Rightarrow} \langle \text{Aozhou shi } X_{\boxed{1}} \text{ de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that } X_{\boxed{1}} \rangle$

$\overset{\pi_3}{\Rightarrow} \langle \text{Aozhou shi yu } X_{\boxed{1}} \text{ you } X_{\boxed{0}} \text{ de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that have } X_{\boxed{0}} \text{ with } X_{\boxed{1}} \rangle$

$\overset{\pi_7}{\Rightarrow} \langle \text{Aozhou shi yu Beihan you } X_{\boxed{0}} \text{ de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that have } X_{\boxed{0}} \text{ with North Korea} \rangle$

$\overset{\pi_9}{\Rightarrow} \langle \text{Aozhou shi yu Beihan you bangjiao de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that have diplomatic relations with North Korea} \rangle$

$\overset{\pi_{10}}{\Rightarrow} \langle \text{Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi,}$
$\text{Australia is one of the few countries that have diplomatic relations with N. K.} \rangle$

Derivation

$\langle S, S \rangle$

$\stackrel{\pi_1}{\Rightarrow} \langle S\ X, S\ X \rangle$

$\stackrel{\pi_1}{\Rightarrow} \langle S\ X_{\boxed{1}}\ X_{\boxed{2}}, S\ X_{\boxed{1}}\ X_{\boxed{2}} \rangle$

$\stackrel{\pi_2}{\Rightarrow} \langle X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}, X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}} \rangle$

$\stackrel{\pi_6}{\Rightarrow} \langle \text{Aozhou } X_{\boxed{1}}\ X_{\boxed{2}}, \text{Australia } X_{\boxed{1}}\ X_{\boxed{2}} \rangle$

$\stackrel{\pi_8}{\Rightarrow} \langle \text{Aozhou shi X, Australia is X} \rangle$

$\stackrel{\pi_5}{\Rightarrow} \langle \text{Aozhou shi X zhiyi, Australia is one of X} \rangle$

$\stackrel{\pi_4}{\Rightarrow} \langle \text{Aozhou shi } X_{\boxed{1}} \text{ de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that } X_{\boxed{1}} \rangle$

$\stackrel{\pi_3}{\Rightarrow} \langle \text{Aozhou shi yu } X_{\boxed{1}} \text{ you } X_{\boxed{0}} \text{ de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that have } X_{\boxed{0}} \text{ with } X_{\boxed{1}} \rangle$

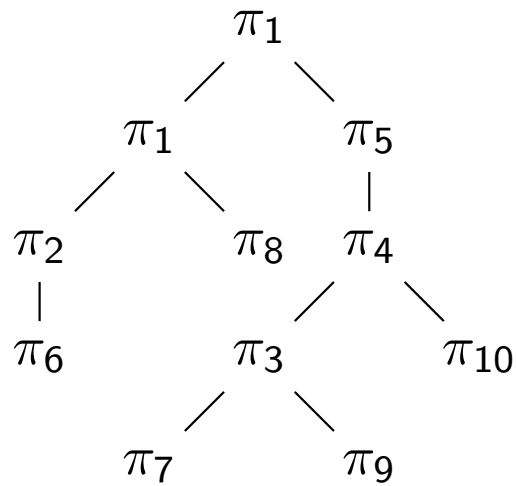
$\stackrel{\pi_7}{\Rightarrow} \langle \text{Aozhou shi yu Beihan you } X_{\boxed{0}} \text{ de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that have } X_{\boxed{0}} \text{ with North Korea} \rangle$

$\stackrel{\pi_9}{\Rightarrow} \langle \text{Aozhou shi yu Beihan you bangjiao de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that have diplomatic relations with North Korea} \rangle$

$\stackrel{\pi_{10}}{\Rightarrow} \langle \text{Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi,}$
$\text{Australia is one of the few countries that have diplomatic relations with N. K.} \rangle$

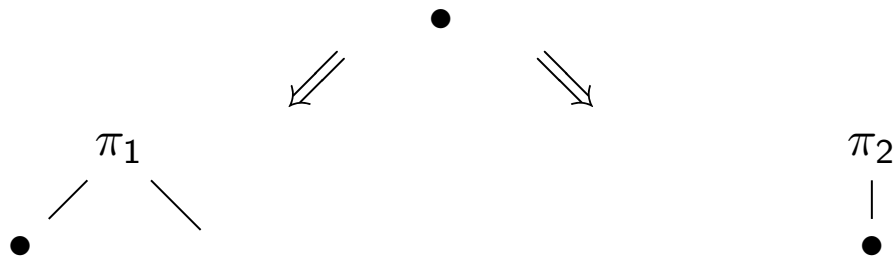
Inducing a Model

Derivation Tree



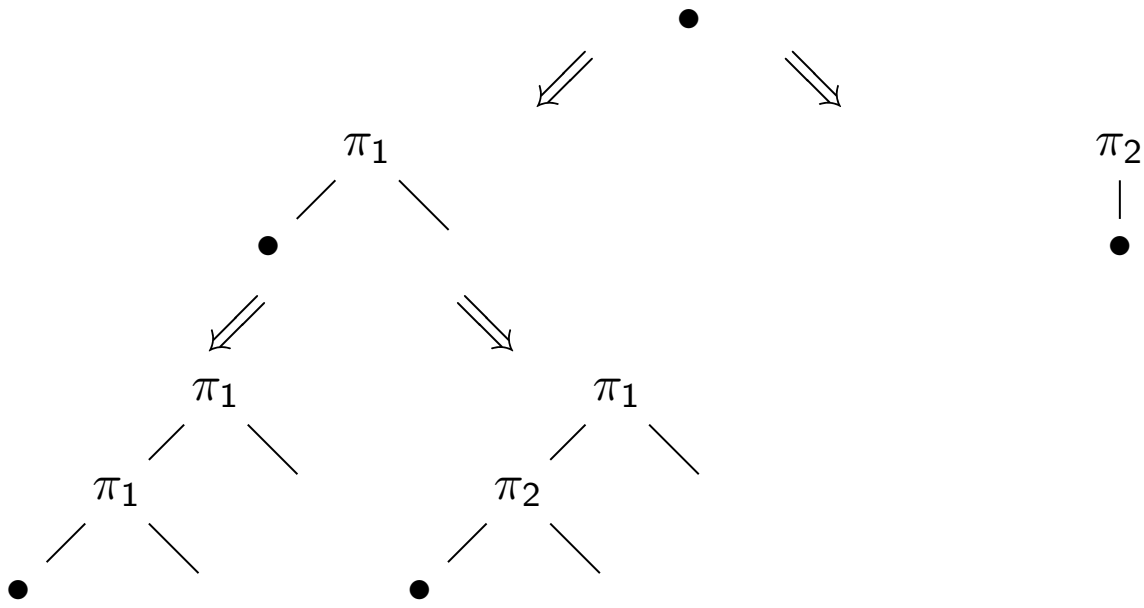
Inducing a Model

Decision Tree



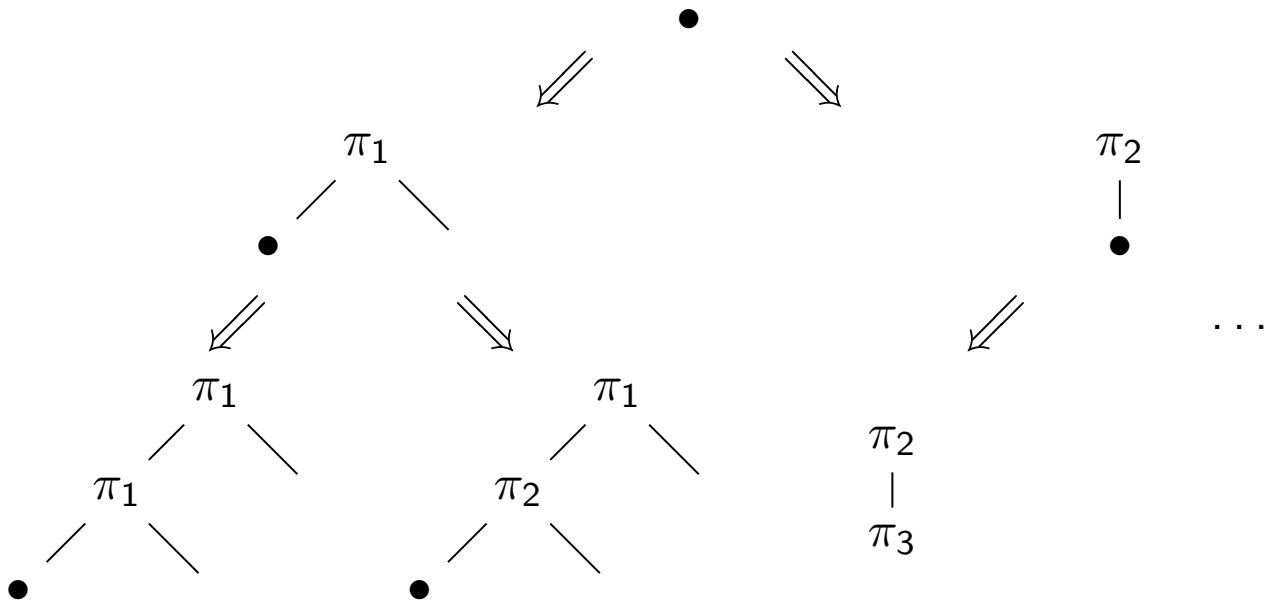
Inducing a Model

Decision Tree



Inducing a Model

Decision Tree



Inducing a Model

Assignments

$$\left.\begin{array}{r}\text{grammar } G \\ \text{assignment } p'\end{array}\right\} \text{“distribution”} \;\; p'(e, d, f)$$

$$\text{model } \mathcal{M}_G = \{p \in \mathcal{M}(E, F) \mid \exists p' \colon p(e, f) = \textstyle\sum_d p'(e, d, f)\}$$

Proper and Consistent Assignments

Let $q \in [0, 1]$.

$$\begin{array}{lll} \pi_1\colon & \mathsf{S} \to \langle \mathsf{S}\ \mathsf{S}, \mathsf{S}\ \mathsf{S} \rangle & \# \ q \\ \pi_2\colon & \mathsf{S} \to \langle \varepsilon, \varepsilon \rangle & \# \ 1 - q \end{array}$$

Proper and Consistent Assignments

Let $q \in [0, 1]$.

$$\begin{aligned} \pi_1\colon &\quad \mathsf{S} \to \langle \mathsf{S}\ \mathsf{S},\ \mathsf{S}\ \mathsf{S} \rangle && \#\ q \\ \pi_2\colon &\quad \mathsf{S} \to \langle \varepsilon,\ \varepsilon \rangle && \#\ 1 - q \end{aligned}$$

Consider $\sum_d p(\varepsilon, d, \varepsilon)$. This is the least solution to

$$Z = q \cdot Z \cdot Z + (1 - q)\ .$$

Proper and Consistent Assignments

Let $q \in [0, 1]$.

$$\begin{aligned} \pi_1\colon &\quad \mathsf{S} \to \langle \mathsf{S}\ \mathsf{S}, \mathsf{S}\ \mathsf{S} \rangle && \# \ q \\ \pi_2\colon &\quad \mathsf{S} \to \langle \varepsilon, \varepsilon \rangle && \# \ 1 - q \end{aligned}$$

Consider $\sum_d p(\varepsilon, d, \varepsilon)$. This is the least solution to

$$Z = q \cdot Z \cdot Z + (1 - q)\ .$$

Reordering gives $(q \neq 0)$

$$0 = Z^2 - \frac{1}{q} \cdot Z + \frac{1 - q}{q}$$

Proper and Consistent Assignments

Let $q \in [0, 1]$.

$$\begin{aligned} \pi_1\colon &\quad \mathsf{S} \to \langle \mathsf{S}\ \mathsf{S},\ \mathsf{S}\ \mathsf{S} \rangle && \#\ q \\ \pi_2\colon &\quad \mathsf{S} \to \langle \varepsilon,\ \varepsilon \rangle && \#\ 1 - q \end{aligned}$$

Consider $\sum_d p(\varepsilon, d, \varepsilon)$. This is the least solution to

$$Z = q \cdot Z \cdot Z + (1 - q)\ .$$

Reordering gives $(q \neq 0)$

$$0 = Z^2 - \frac{1}{q} \cdot Z + \frac{1 - q}{q}$$

Then

$$Z_1 = 1 \qquad\qquad Z_2 = \frac{1}{q} - 1$$

Proper and Consistent Assignments

Let $q \in [0, 1]$.

$$\begin{aligned} \pi_1\colon &\quad \mathsf{S} \to \langle \mathsf{S}\ \mathsf{S}, \mathsf{S}\ \mathsf{S} \rangle && \# \ q \\ \pi_2\colon &\quad \mathsf{S} \to \langle \varepsilon, \varepsilon \rangle && \# \ 1 - q \end{aligned}$$

Consider $\sum_d p(\varepsilon, d, \varepsilon)$. This is the least solution to

$$Z = q \cdot Z \cdot Z + (1 - q)\ .$$

Reordering gives $(q \neq 0)$

$$0 = Z^2 - \frac{1}{q} \cdot Z + \frac{1 - q}{q}$$

Then

$$Z_1 = 1 \qquad\qquad Z_2 = \frac{1}{q} - 1$$

Least solution:

$$Z = \begin{cases} 1 & \text{if } 0 \leq q \leq 0.5, \\ \frac{1}{q} - 1 & \text{otherwise.} \end{cases}$$

Training: Rule Extraction

Word-aligned Training Pair

⟨30 duonianlai de youhao hezuo, over 30 years of friendly cooperation⟩

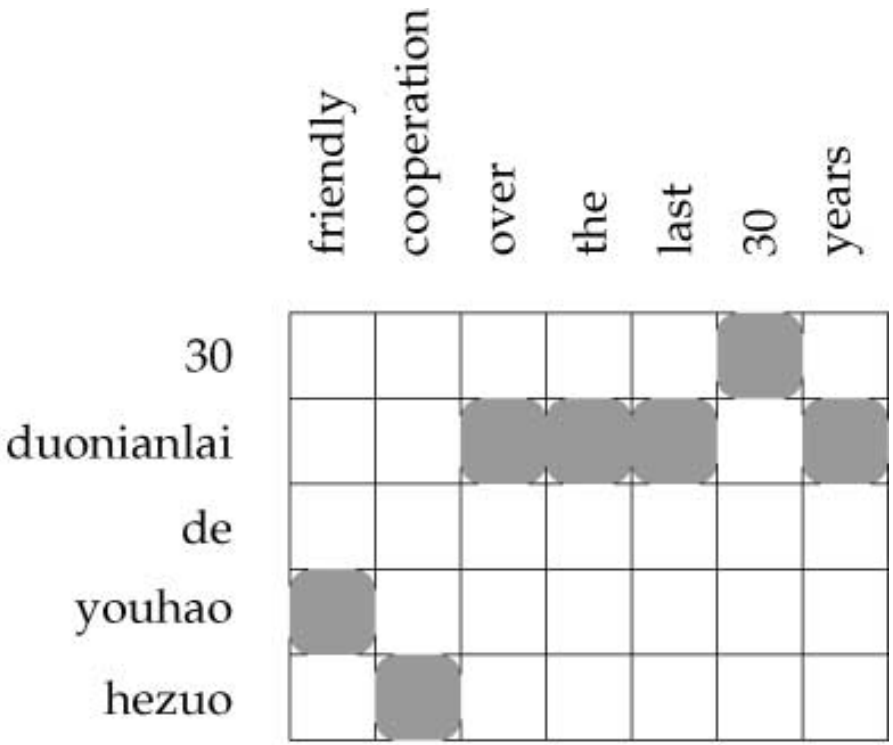


Figure 2 of (Chiang 2007)

Training: Rule Extraction

Initial Rules

X → ⟨30 duonianlai de youhao hezuo, over 30 years of friendly cooperation⟩
X → ⟨duonianlai, over the last 30 years⟩
X → ⟨youhao hezuo, friendly cooperation⟩
X → ⟨30, 30⟩
X → ⟨youhao, friendly⟩
X → ⟨hezuo, cooperation⟩

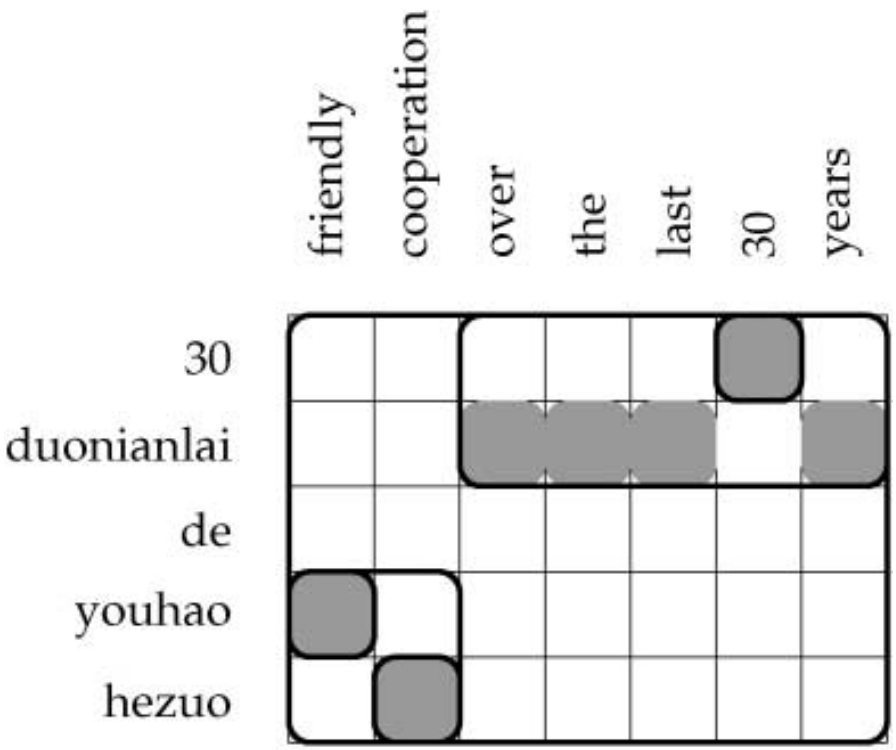


Figure 2 of (Chiang 2007)

Training: Rule Extraction

Example Rule

$X \rightarrow \langle X_{[1]}$ duonianlai de $X_{[2]}$, $X_{[2]}$ over the last years $X_{[1]} \rangle$

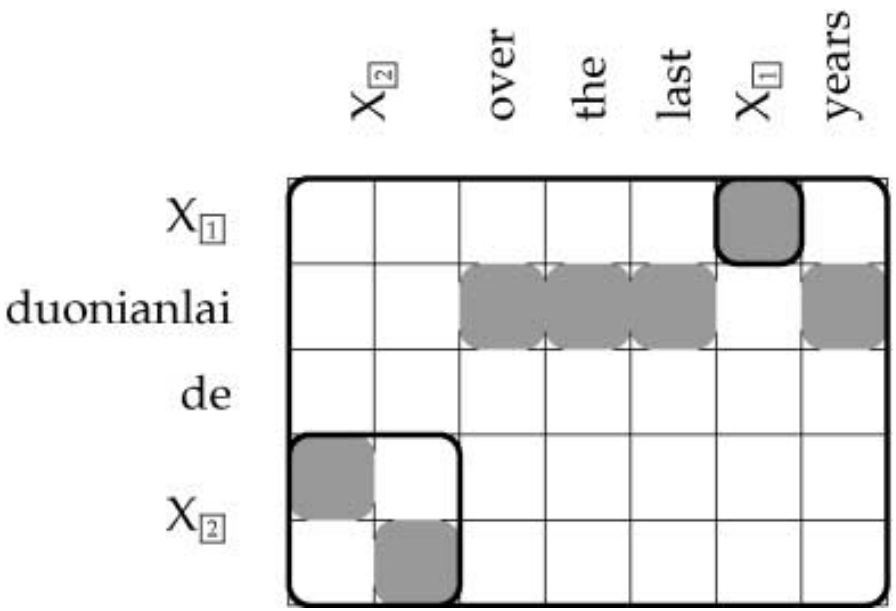


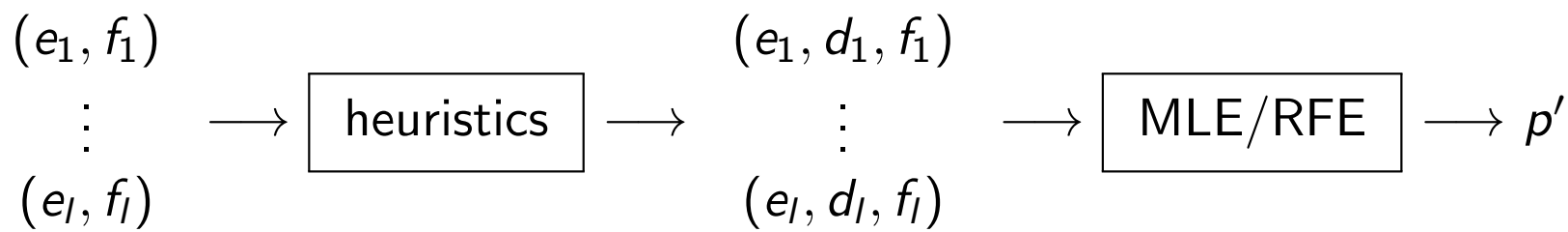
Figure 2 of (Chiang 2007)

Training: Maximum-Likelihood Estimation

- training data: sequence $\mathcal{D} = (e_1, f_1), \ldots, (e_l, f_l)$
- model: $\mathcal{M}_G = \{p \in \mathcal{M}(E, F) \mid \exists p' : p(e, f) = \sum_d p'(e, d, f)\}$
- likelihood: $P(\mathcal{D} \mid p') = \prod_j \sum_d p'(e_j, d, f_j)$
- use EM algorithm

Training: Reality

Chiang (2007), DeNeefe and Knight (2009):



Outline

Feature Selection

Hidden variable: derivation of (e, f) in G

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg
- $h_3(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg
- $h_3(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model
- $h_4(e, f) = -|e|$
  length preference

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg
- $h_3(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model
- $h_4(e, f) = -|e|$
  length preference

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg
- $h_3(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model
- $h_4(e, f) = -|e|$
  length preference

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$
- $h_1(e, d, f) = \log p(e)$

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg
- $h_3(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model
- $h_4(e, f) = -|e|$
  length preference

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$
- $h_1(e, d, f) = \log p(e)$
- $h_2(e, d, f) = \log p_G(e, d \mid f)$

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg
- $h_3(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model
- $h_4(e, f) = -|e|$
  length preference

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$
- $h_1(e, d, f) = \log p(e)$
- $h_2(e, d, f) = \log p_G(e, d \mid f)$
- $h_3(e, d, f) = \log p_G(f, d \mid e)$

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg
- $h_3(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model
- $h_4(e, f) = -|e|$
  length preference

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$
- $h_1(e, d, f) = \log p(e)$
- $h_2(e, d, f) = \log p_G(e, d \mid f)$
- $h_3(e, d, f) = \log p_G(f, d \mid e)$
- $h_4(e, d, f) = -|e|$

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg
- $h_3(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model
- $h_4(e, f) = -|e|$
  length preference

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$
- $h_1(e, d, f) = \log p(e)$
- $h_2(e, d, f) = \log p_G(e, d \mid f)$
- $h_3(e, d, f) = \log p_G(f, d \mid e)$
- $h_4(e, d, f) = -|e|$
- $h_5(e, d, f) =$ lexical weights

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg
- $h_3(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model
- $h_4(e, f) = -|e|$
  length preference

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$
- $h_1(e, d, f) = \log p(e)$
- $h_2(e, d, f) = \log p_G(e, d \mid f)$
- $h_3(e, d, f) = \log p_G(f, d \mid e)$
- $h_4(e, d, f) = -|e|$
- $h_5(e, d, f) =$ lexical weights
- $h_6(e, d, f) = -\#_{\text{ex}} d$

Feature Selection

Hidden variable: derivation of (e, f) in G

- log-linear model $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log p(e)$
  n-gram language model
- $h_2(e, f) = \log \sum_d p_G(e, d, f)$
  synchronous cfg
- $h_3(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model
- $h_4(e, f) = -|e|$
  length preference

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$
- $h_1(e, d, f) = \log p(e)$
- $h_2(e, d, f) = \log p_G(e, d \mid f)$
- $h_3(e, d, f) = \log p_G(f, d \mid e)$
- $h_4(e, d, f) = -|e|$
- $h_5(e, d, f) =$ lexical weights
- $h_6(e, d, f) = -\#_{\text{ex}} d$
- $h_7(e, d, f) = -\#_{\text{gl}} d$

Decoding

$$\hat{e} = \text{argmax}_e \; p(e \mid f)$$

Decoding

$$\begin{aligned}\hat{e} &= \text{argmax}_e \; p(e \mid f) \\ &= \text{argmax}_e \sum_d p(e, d \mid f)\end{aligned}$$

Decoding

$$
\begin{aligned}
\hat{e} &= \text{argmax}_e \; p(e \mid f) \\
&= \text{argmax}_e \; \sum_d p(e, d \mid f) \\
&= \text{argmax}_e \; \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)}
\end{aligned}
$$

Decoding

$$
\begin{aligned}
\hat{e} &= \text{argmax}_e \; p(e \mid f) \\
&= \text{argmax}_e \; \sum_d p(e, d \mid f) \\
&= \text{argmax}_e \; \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)} \\
&= \text{argmax}_e \; \sum_d \exp \sum_i \lambda_i \cdot h_i(e, d, f)
\end{aligned}
$$

Decoding

$$
\begin{aligned}
\hat{e} &= \text{argmax}_e \; p(e \mid f) \\
&= \text{argmax}_e \sum_d p(e, d \mid f) \\
&= \text{argmax}_e \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)} \\
&= \text{argmax}_e \sum_d \exp \sum_i \lambda_i \cdot h_i(e, d, f)
\end{aligned}
$$

NP-hard! (Casacuberta and Higuera 2000; Li, Eisner, and Khudanpur 2009)

Decoding

$$\begin{aligned}
\hat{e} &= \text{argmax}_e \; p(e \mid f) \\
&= \text{argmax}_e \sum_d p(e, d \mid f) \\
&= \text{argmax}_e \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)} \\
&= \text{argmax}_e \sum_d \exp \sum_i \lambda_i \cdot h_i(e, d, f)
\end{aligned}$$

NP-hard! (Casacuberta and Higuera 2000; Li, Eisner, and Khudanpur 2009)

$$\approx \text{argmax}_e \sum_{d \in D_{\text{fin}}} \exp \sum_i \lambda_i \cdot h_i(e, d, f)$$

Decoding

$$\hat{e} = \text{argmax}_e \; p(e \mid f)$$

$$= \text{argmax}_e \sum_d p(e, d \mid f)$$

$$= \text{argmax}_e \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)}$$

$$= \text{argmax}_e \sum_d \exp \sum_i \lambda_i \cdot h_i(e, d, f)$$

NP-hard! (Casacuberta and Higuera 2000; Li, Eisner, and Khudanpur 2009)

$$\approx \text{argmax}_e \sum_{d \in D_{\text{fin}}} \exp \sum_i \lambda_i \cdot h_i(e, d, f)$$

where $D_{\text{fin}} = \text{argmax}_d^{(n)} \exp \sum_i \lambda_i \cdot h_i(\text{e}(d), d, f)$

Decoding

$$\hat{e} = \text{argmax}_e \; p(e \mid f)$$

$$= \text{argmax}_e \sum_d p(e, d \mid f)$$

$$= \text{argmax}_e \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)}$$

$$= \text{argmax}_e \sum_d \exp \sum_i \lambda_i \cdot h_i(e, d, f)$$

NP-hard! (Casacuberta and Higuera 2000; Li, Eisner, and Khudanpur 2009)

$$\approx \text{argmax}_e \sum_{d \in D_{\text{fin}}} \exp \sum_i \lambda_i \cdot h_i(e, d, f)$$

where $D_{\text{fin}} = \text{argmax}_d^{(n)} \exp \sum_i \lambda_i \cdot h_i(\text{e}(d), d, f)$

Cube Pruning (Chiang 2007)

Banerjee, Satanjeev and Alon Lavie (2005). "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments". In: Proceedings of the ACL 2005 Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization.

Berger, Adam L., Vincent J. Della Pietra, and Stephen A. Della Pietra (1996). "A maximum entropy approach to natural language processing". In: Comp. Ling. 22 (1), pp. 39–71. ISSN: 0891-2017. URL: http://portal.acm.org/citation.cfm?id=234285.234289.

Brown, Peter F. et al. (1993). "The mathematics of statistical machine translation: parameter estimation". In: Comp. Ling. 19 (2), pp. 263–311. ISSN: 0891-2017. URL: http://portal.acm.org/citation.cfm?id=972470.972474.

Casacuberta, Francisco and Colin de la Higuera (2000). "Computational Complexity of Problems on Probabilistic Grammars and Transducers". In: LNCS.

Casella, George and Roger L. Berger (2002). Statistical Inference. Duxbury Advances Series. Duxbury.

Chiang, David (2007). "Hierarchical Phrase-Based Translation". In: Comp. Ling. 33.2, pp. 201–228. ISSN: 0891-2017. DOI: http://dx.doi.org/10.1162/coli.2007.33.2.201.

DeNeefe, Steve and Kevin Knight (2009). "Synchronous tree adjoining machine translation". In: EMNLP '09: Proceedings of the 2009

Conference on Empirical Methods in Natural Language Processing. Singapore: Association for Computational Linguistics, pp. 727–736. ISBN: 978-1-932432-62-6.

Li, Zhifei, Jason Eisner, and Sanjeev Khudanpur (2009). "Variational decoding for statistical machine translation". In: ACL-IJCNLP '09: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2. Suntec, Singapore: Association for Computational Linguistics, pp. 593–601. ISBN: 978-1-932432-46-6.

Och, Franz Josef (2003). "Minimum Error Rate Training in Statistical Machine Translation". In: ACL, pp. 160–167.

Och, Franz Josef and Hermann Ney (2002). "Discriminative Training and Maximum Entropy Models for Statistical Machine Translation". In: ACL, pp. 295–302.

Papineni, Kishore et al. (2002). "BLEU: a method for automatic evaluation of machine translation". In: ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Philadelphia, Pennsylvania: Association for Computational Linguistics, pp. 311–318. DOI: http://dx.doi.org/10.3115/1073083.1073135.

Smith, David A. and Jason Eisner (2006). "Minimum-Risk Annealing for Training Log-Linear Models". In: Proceedings of the International Conference on Computational Linguistics and the Association for

Computational Linguistics (COLING-ACL), Companion Volume. Sydney, pp. 787–794. URL: http://cs.jhu.edu/~jason/papers/#acl06-risk.

Snover, Matthew et al. (2006). "A Study of Translation Edit Rate with Targeted Human Annotation". In: Proceedings of Association for Machine Translation in the Americas.