

翻訳モデルと翻訳エンジン

内山将夫@NICT

mutiyama@nict.go.jp

これまでのまとめ

- MT の性能評価
- 言語モデル
- 単語対応のモデル化
- 対訳コーパスの作成

これからの目次

- 翻訳モデルと翻訳エンジンの関係
- 句(フレーズ)に基づく翻訳モデル
- 句表(フレーズテーブル)の作成
- 対数線形モデルとデコーダー
- 誤り率最小化訓練 (MERT, minimum error rate training)

翻訳モデルと翻訳エンジンの関係

- 翻訳エンジンというのは，入力文 f を受け取り，それを翻訳して，出力文 \hat{e} とするものである .
- そのときには，何らかの探索アルゴリズムを利用して，最適スコアの出力文 \hat{e} を得る .
- そのスコアの付け方 g を翻訳モデルという

$$\hat{e} = \arg \max_e g(e, f)$$

- \hat{e} は出力文
- $\arg \max_e$ は最適解の探索を示す
- $g(e, f)$ が翻訳モデル
- 翻訳エンジンは，翻訳モデルに従って，最適な翻訳を得るためのものである .

探索アルゴリズムと翻訳モデルの密接な関係

1. 出力候補の数はとても多い .

入力 $f = f_1 f_2 \dots f_m$ が , それぞれ , 3 単語の対訳単語をもつとすると , それだけで , 3^m の組み合わせがある . さらに順番が入れかわるので , $m!3^m$ の組み合わせがある . これは , 単語が 1 対 1 に翻訳された場合でもである .

m	1	3	5	7	10
$m!3^m$	3	162	29160	1 千万	2000 億

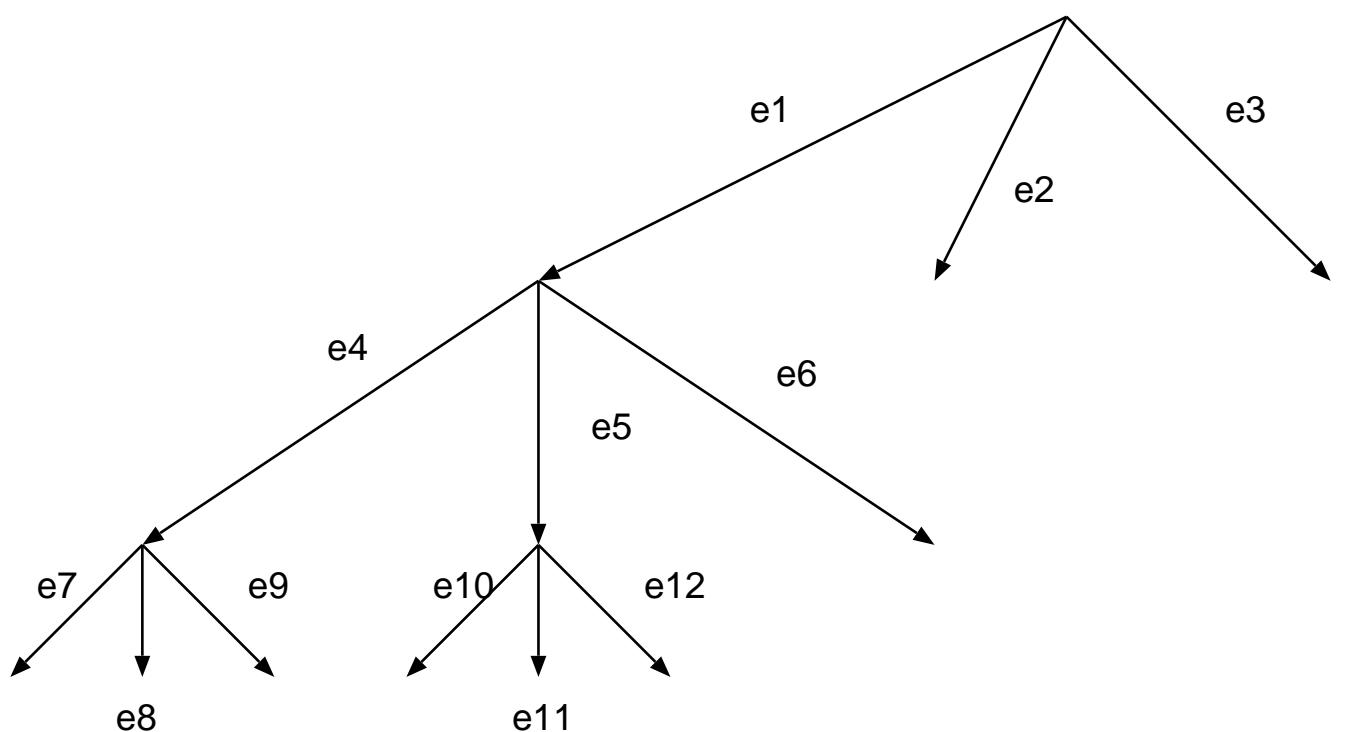
のように単語数 m に応じて , すぐに大きくなる .

→ 全部を列挙することは不可能

探索アルゴリズムと翻訳モデルの密接な関係

2. 探索は翻訳モデルにガイドされる

少しずつ出力文をつくっていき，その都度，翻訳モデルにより割当てられるスコアの小さい候補を消していく



良い翻訳モデルならば良いスコアを与えることができるの，いらない候補を早く消すことができる．また，スコアの計算方法と探索法とには，密接な関係がある．
→ 実際上スコアの計算法は，探索方法により決まる

ただし，たとえば，混合整数計画法により，スコアを定義することも可能なので，そのうちに，探索法は，MTの研究の枠外になるかもしれない．

統計的翻訳モデルの歴史概観

1. 単語単位の翻訳モデル (1980年代後半～1990年代)
2. 句単位の翻訳モデル (1990年代後半～)
3. 構造に基づく翻訳モデル (2000年代半ば～)

2や3の始まりは、1に少し遅れた時期だが、研究が盛んになったのは、それぞれ、2000年～、2005年～くらいである。現在は、構造に基づく翻訳モデルに研究の主流は移ってきてている。

しかし、句単位の翻訳は、簡単でかつ効率的で、構造に基づく翻訳の基礎となるものなので、ここでは、句単位の翻訳モデルを紹介する。