16. 対数線形モデルの導入

内山将夫@NICT
mutiyama@nict.go.jp

翻訳候補へのスコア付け

入力文を $\mathbf{f}$
翻訳候補を $\mathbf{e}$
$\mathbf{e}$ のスコアを $S_{\mathbf{f}}(\mathbf{e})$ とする．
翻訳システム (デコーダー) は

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}} S_{\mathbf{f}}(\mathbf{e})$$

なる $\hat{\mathbf{e}}$ を出力する．
$\arg\max$ の部分は，デコーダーが最大スコアであるような $S_{\mathbf{f}}(\mathbf{e})$ を探す部分である．
以下では，スコア付けの方法について述べる．

確率に基づくスコア付け

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}} P(\mathbf{e}|\mathbf{f})$$

つまり，入力$\mathbf{f}$に対して，最大確率となる$\hat{\mathbf{e}}$を出力するのが基本である．その確率の定式化として

1. ベイズの定理に基づく方法
2. 対数線形モデルに基づく方法

の2通りがある．
2は1の一般化であり，現状では，多くのSMTシステムがこの方法を利用している．まず，1から説明する．

ベイズの定理に基づく方法

$$\begin{aligned}\hat{\mathbf{e}} &= \arg\max_{\mathbf{e}} P(\mathbf{e}|\mathbf{f}) \\ &= \arg\max_{\mathbf{e}} P(\mathbf{f}|\mathbf{e})P(\mathbf{e})\end{aligned}$$

つまり $P(\mathbf{e}|\mathbf{f})$ のかわりに $P(\mathbf{f}|\mathbf{e})P(\mathbf{e})$ を計算する．ここで

- $P(\mathbf{e})$ = 言語モデルによる $\mathbf{e}$ の確率．つまり，生成された $\mathbf{e}$ のつくりの良さを確率で表現したもの
- $P(\mathbf{f}|\mathbf{e})$ = $\mathbf{e}$ を条件としたときの $\mathbf{f}$ の確率．$\mathbf{f}$ の翻訳としての良さを確率で表現したもの

これについては,「初歩の確率」のときに少し言及してある.

ベイズの定理と現実との乖離

ベイズの定理の式自体は，正しい式である．しかし，現実には，$P(\cdot)$ の推定に様々な誤りが含まれるため，ベイズの定理に基づく方法は，現実には最適な方法ではない．
$P(\cdot)$ の推定における誤りには，以下のものがある．

- モデルの誤り
- パラメタ推定の誤り

モデルの誤りとは，対象の確率モデルを作るときに，そのモデル自体に誤りがあるため，たとえそのモデルに従って完全なパラメタ推定ができたとしても，対象を正確に表現できないことを言う．実際には，言語を正確に確率モデルとして表現したものは存在しないので，全てのモデルには，程度の大小があったとしても，モデルの誤りがある．
パラメタ推定の誤りとは，与えられたモデルに従って，与えられたデータからモデルのパラメタを推定したときに，データの性質やパラメタ推定法等により，パラメタ推定が上手くいかないことである．パラメタ推定の誤りも必ず存在する．

ベイズの定理に基づく方法でできないこと

1. 翻訳モデルと言語モデルの重みを変えることができない．しかし
$$P(\mathbf{f}|\mathbf{e})^{\lambda_1}P(\mathbf{e})^{\lambda_2}$$
のように重み付けをすることにより，翻訳精度が高くなることがわかっている．
2. 色々な情報を入れることができない．たとえば，$P(\mathbf{f}|\mathbf{e})$ だけでなく，$P(\mathbf{e}|\mathbf{f})$ を追加して，
$$P(\mathbf{f}|\mathbf{e})^{\lambda_1}P(\mathbf{e})^{\lambda_2}P(\mathbf{e}|\mathbf{f})^{\lambda_3}$$
のようにしたい．

上記のようなことをするためには，対数線形モデルが便利である．

ベイズの定理から対数線形モデルに

$$P(\mathbf{e}|\mathbf{f}) = \frac{P(\mathbf{f}|\mathbf{e})P(\mathbf{e})}{P(\mathbf{f})} = \frac{P(\mathbf{f}|\mathbf{e})P(\mathbf{e})}{\Sigma_{\mathbf{e}'} P(\mathbf{f}|\mathbf{e}')P(\mathbf{e}')}$$

$$\begin{aligned}
P(\mathbf{f}|\mathbf{e})P(\mathbf{e}) &= \exp(\log P(\mathbf{f}|\mathbf{e}) + \log P(\mathbf{e})) \\
&= \exp(\lambda_1 \log P(\mathbf{f}|\mathbf{e}) + \lambda_2 \log P(\mathbf{e})) \\
&= \exp(\lambda_1 h_1(\mathbf{f},\mathbf{e}) + \lambda_2 h_2(\mathbf{f},\mathbf{e})) \\
&= \exp(\sum_{i=1}^{2} \lambda_i h_i(\mathbf{f},\mathbf{e}))
\end{aligned}$$

ただし，

$$\begin{aligned}
\lambda_1 &= 1 \\
\lambda_2 &= 1 \\
h_1(\mathbf{f},\mathbf{e}) &= \log P(\mathbf{f}|\mathbf{e}) \\
h_2(\mathbf{f},\mathbf{e}) &= \log P(\mathbf{e})
\end{aligned}$$

よって，

$$P(\mathbf{e}|\mathbf{f}) = \frac{\exp(\Sigma_{i=1}^{2} \lambda_i h_i(\mathbf{f},\mathbf{e}))}{\Sigma_{\mathbf{e}'} \exp(\Sigma_{i=1}^{2} \lambda_i h_i(\mathbf{f},\mathbf{e}'))}$$

一般化

M 個の関数 $h_i(\mathbf{e}, \mathbf{f})$
M 個の重み λ_i について

$$\begin{aligned} \hat{\mathbf{e}} &= \arg\max_{\mathbf{e}} P(\mathbf{e}|\mathbf{f}) \\ &= \arg\max_{\mathbf{e}} \frac{\exp(\Sigma_{i=1}^{M} \lambda_i h_i(\mathbf{f}, \mathbf{e}))}{\Sigma_{\mathbf{e}'} \exp(\Sigma_{i=1}^{M} \lambda_i h_i(\mathbf{f}, \mathbf{e}'))} \\ &= \arg\max_{\mathbf{e}} \sum_{i=1}^{M} \lambda_i h_i(\mathbf{f}, \mathbf{e}) \end{aligned} \quad (1)$$

ようするに，各関数 $h_i(\mathbf{e}, \mathbf{f})$ の重み付きの和をもってスコアとする．
この関数を素性という．
対数線形モデルを利用することにより，様々な素性をスコアとして考慮できる．この素性は良く考えて決める．
たとえば，フレーズテーブルのスコアが素性のスコアとなる．
重みは，後述の手法により，自動的に決める．