条件付き正規化最尤分布とその評価尺度について

北海道大・情報科学　廣瀬 善大

1 概要

条件付き正規化最尤分布（Conditional Normalized Maximum Likelihood distribution, CNML）とは，正規化最尤分布（Normalized Maximum Likelihood distribution, NML）と呼ばれている確率分布を条件付き確率分布として一般化したものである．NML はリグレットと呼ばれる損失関数のミニマックスを達成する確率分布として定義される．一方，CNML はリグレットと類似の損失関数に関してミニマックスを達成する条件付き分布として定義されたものである．本発表では，過去の観測値をもとに将来の観測値の確率分布を推定する分布予測を考える．

2 条件付きリグレットと条件付き正規化最尤分布

標本空間 $\mathcal{X}$ とその上の統計モデル $\mathcal{M} = \{P(\cdot;\theta)|\theta \in \Theta\}$ を考える．ただし，Θ はパラメータ空間である．リグレット損失とは，確率分布 P と $x \in \mathcal{X}$ に対して $\mathrm{Reg}(P,x) = -\log P(x) + \log P(x;\hat{\theta}(x))$ と定義される損失関数である．ただし，$\hat{\theta}(\cdot)$ はパラメータ θ の最尤推定量である．リグレットのミニマックスを達成する確率分布が存在する場合，その確率分布が NML である [2]．$\mathcal{X} \times \mathcal{X}$ 上のモデル $\widetilde{\mathcal{M}} = \{P(\cdot,\cdot;\theta)|\theta \in \Theta\}$ を考えることで条件付き分布に対する損失関数をリグレットと同様に定義することができる．つまり，条件付き確率分布 P と $x, y \in \mathcal{X}$ に対して

$$\mathrm{CReg}(P,y,x) = -\log P(y|x) + \log P(y|x;\hat{\theta}(x,y))$$

という損失関数を定義することができる [1]．ここで $\hat{\theta}(\cdot,\cdot)$ はモデル $\widetilde{\mathcal{M}}$ におけるパラメータ θ の最尤推定量である．損失関数 CReg を条件付きリグレットと呼ぶことにすると，CNML は $x \in \mathcal{X}$ ごとに条件付きリグレットのミニマックスを達成する条件付き分布として定義される：$P_{\mathrm{CNML}}(\cdot|x) := \arg\min_P \max_{y\in\mathcal{X}} \mathrm{CReg}(P,y,x)$, $x \in \mathcal{X}$. ただし，最小化に現れる $P(\cdot|x)$ はすべての確率分布を候補とし，モデル $\widetilde{\mathcal{M}}$ から得られる条件付き分布とは限らない．

本発表では，分布予測において CNML を考える．特に，カルバック–ライブラー情報量以外の評価尺度のもとでの CNML の挙動や，CNML に関連する評価尺度そのものの性質について報告する．

参考文献

[1] Grünwald (2007). *The Minimum Description Length Principle*, MIT Press.

[2] Shtarkov (1987). Universal Sequential Coding of Single Messages, *Problems of Information Transmission*, **23**, 3–17.