# Rate Distortion Theory and Information Bottleneck Method

Julien Wolf

17. September 2008

# Inhalt

# Problemstellung

- Information ist teuer! Daher: »Was ist relevante Information«?
- Idee: zulässige Veränderung der Information quantifizieren
- innerhalb dieser Grenze (verlustbehaftet) komprimieren
- Übertragung der Information in ein vereinfachtes Alphabet
  $\rightsquigarrow$ "lossy source compression" $\rightsquigarrow$ *rate distortion theory*

# Definitionen

- $X$ (endlicher) Signalraum
- $\tilde{X}$ "codebook" / komprimiertes Signal
- stochastische Abbildung $X \to \tilde{X}$ durch $p(\tilde{x}|x)$
  $\rightsquigarrow$ Partitionierung von $X$ durch $p(\tilde{x}|x)$, mit

$$p(\tilde{x}) = \sum_x p(x)p(\tilde{x}|x) \tag{1}$$

- Maß für die Kompressionsrate: $I(X;\tilde{X})$
- distortion function $d\colon X \times \tilde{X} \to \mathbb{R}_0^+$
- Verzerrung bei der Partitionierung von $X$:

$$\langle d(x,\tilde{x})\rangle_{p(x,\tilde{x})} = \sum_{x\in X}\sum_{\tilde{x}\in\tilde{X}} p(x,\tilde{x})d(x,\tilde{x}) \tag{2}$$

# Rate distortion function

## Definition (rate distortion function)

Es gibt für jede gegebene Verzerrung $D$ eine Untergrenze für die Kompressionsrate:

$$R(D) \equiv \min_{\{p(\tilde{x}|x):\ \langle d(x,\tilde{x})\rangle \leq D\}} I(X;\tilde{X}) \tag{3}$$

Mit Lagrange-Multiplikator $\beta$:

$$\mathcal{F}[p(\tilde{x}|x)] = I(X;\tilde{X}) + \beta \left\langle d(x,\tilde{x}) \right\rangle_{p(x,\tilde{x})} \tag{4}$$

# Eigenschaften von $R(D)$

## Theorem

*Lösung des Minimierungsproblems:*

$$p(\tilde{x}|x) = \frac{p(\tilde{x})}{Z(x,\beta)} \exp\left(-\beta d(x,\tilde{x})\right) \quad , \tag{5}$$

*wobei* $Z(x,\beta)$ *eine Zustandssumme ist. Ausserdem ist* $\beta > 0$ *und es gilt:*

$$\frac{\delta R}{\delta D} = -\beta \tag{6}$$

# Beweis

$$\mathcal{F}[p(\tilde{x}|x)] = \sum_x \sum_{\tilde{x}} p(x)p(\tilde{x}|x) \log \frac{p(\tilde{x}|x)}{p(\tilde{x})} \tag{7}$$
$$+ \beta \sum_x \sum_{\tilde{x}} p(x)p(\tilde{x}|x)p(x)d(x,\tilde{x})$$
$$+ \sum_x \lambda(x) \sum_{\tilde{x}} p(\tilde{x}|x)$$

$$\left.\frac{\delta \mathcal{F}}{\delta p(\tilde{x}|x)}\right|_{x,\tilde{x}} = p(x)\left(\log \frac{p(\tilde{x}|x)}{p(\tilde{x})} + 1 - \frac{1}{p(\tilde{x})} \sum_{x'} p(x')p(\tilde{x}|x') \right. \tag{8}$$
$$\left. + \beta d(x,\tilde{x}) + \frac{\lambda(x)}{p(x)}\right)$$
$$= p(x)\left(\log \frac{p(\tilde{x}|x)}{p(\tilde{x})} + \beta d(x,\tilde{x}) + \log Z(x,\beta)\right) \tag{9}$$

# Beweis

$$\mathcal{F}[p(\tilde{x}|x)] = \sum_x \sum_{\tilde{x}} p(x)p(\tilde{x}|x) \log \frac{p(\tilde{x}|x)}{p(\tilde{x})} + \beta \sum_x \sum_{\tilde{x}} p(x)p(\tilde{x}|x)p(x)d(x,\tilde{x}) + \sum_x \lambda(x) \sum_{\tilde{x}} p(\tilde{x}|x) \tag{7}$$

$$\left.\frac{\delta \mathcal{F}}{\delta p(\tilde{x}|x)}\right|_{x,\tilde{x}} = p(x)\left(\log \frac{p(\tilde{x}|x)}{p(\tilde{x})} + 1 - \frac{1}{p(\tilde{x})} \sum_{x'} p(x')p(\tilde{x}|x') + \beta d(x,\tilde{x}) + \frac{\lambda(x)}{p(x)}\right) \tag{8}$$

$$= p(x)\left(\log \frac{p(\tilde{x}|x)}{p(\tilde{x})} + \beta d(x,\tilde{x}) + \log Z(x,\beta)\right) \tag{9}$$

# Blahut-Arimoto-Algorithmus

## Idee:

Minimierung des Abstandes zwischen zwei konvexen Mengen $A, B \in \mathbb{R}^n$ durch abwechselnde Optimierung

## Lemma

*Gegeben* $p(x,y) = p(x)p(y|x)$, *dann ist* $D_{KL}\big(p(x,y)|p(x)p(y)\big)$ *minimal für die Marginalverteilung*

$$p(y) = \sum_x p(x)p(y|x) \tag{10}$$

# Blahut-Arimoto-Algorithmus

Ist $\mathcal{F}$ minimal, so sind die beiden Gleichungen

$$p(\tilde{x}) = \sum_x p(x)p(\tilde{x}|x) \tag{11}$$

$$p(\tilde{x}|x) = \frac{p(\tilde{x})}{Z(x,\beta)} \exp\big(-\beta d(x,\tilde{x})\big) \tag{12}$$

erfüllt. $\mathcal{F}$ kann nun über die die konvexen Mengen $\{p(\tilde{x})\}$ und $\{p(\tilde{x}|x)\}$ minimiert werden:

$$\min_{p(\tilde{x})} \min_{p(\tilde{x}|x)} \mathcal{F}[p(\tilde{x}); p(\tilde{x}|x)] \tag{13}$$

# Blahut-Arimoto-Algorithmus

Iteration:

$$\begin{cases} p_{t+1}(\tilde{x}) = \sum_x p(x) p_t(\tilde{x}|x) \\ p_t(\tilde{x}|x) = \frac{p_t(\tilde{x})}{Z_t(x,\beta)} \exp\big(-\beta d(x,\tilde{x})\big) \end{cases} \tag{14}$$

Konvergiert zu einem eindeutigen Minimum von $\mathcal{F}$.

# Anwendungen

- »relevant quantization« z.B. für Bilder (JPEG, MPEG)
- perception-based distortion measures

# Motivation

- *Bisher: distortion function* muß definiert werden.
- *Idee:* relevant ist, was $X$ über $Y$ aussagt.

$$I(\tilde{X};Y) \leq I(X;Y) \tag{15}$$

- *Ziel:* $I(\tilde{X};X)$ unter Berücksichtigung von $I(\tilde{X};Y)$ minimieren:

$$\mathcal{L}[p(\tilde{x}|x)] = I(\tilde{X};X) - \beta I(\tilde{X};Y) \tag{16}$$

- Für $\beta = 0$ genau ein Cluster
- Für $\beta \to \infty$ keine Kompression

# Eigenschaften

Lösung des Minimierungsproblems:

$$p(\tilde{x}|x) = \frac{p(\tilde{x})}{Z(x,\beta)} \exp\left(-\beta \sum_y p(y|x) \log \frac{p(y|x)}{p(y|\tilde{x})}\right) \tag{17}$$

mit

$$p(y|\tilde{x}) = \frac{1}{p(\tilde{x})} \sum_x p(y|x) p(\tilde{x}|x) p(x) \tag{18}$$

und

$$\frac{\delta I(\tilde{X};Y)}{\delta I(\tilde{X};X)} = \beta^{-1} > 0 \tag{19}$$

# Iterative Berechnung

Ist $\mathcal{F}$ minimal, so gelten die drei Gleichungen

$$p(y|\tilde{x}) = \sum_x p(y|x)p(x|\tilde{x}) \tag{20}$$

$$p(\tilde{x}) = \sum_x p(\tilde{x}|x)p(x) \tag{21}$$

$$p(\tilde{x}|x) = \frac{p(\tilde{x})}{Z(x,\beta)} \exp\left(-\beta\, D_{KL}[p(y|x) \| p(y|\tilde{x})]\right) \tag{22}$$

Analog zu Blahut-Arimoto:

$$\min_{p(y|\tilde{x})} \min_{p(\tilde{x})} \min_{p(\tilde{x}|x)} \mathcal{F}[p(\tilde{x}|x); p(\tilde{x}); p(y|\tilde{x})] \tag{23}$$

# Iterative Berechnung

Iteration:

$$\begin{cases} p_t(\tilde{x}|x) = \frac{p_t(\tilde{x})}{Z_t(x,\beta)} \exp\left(-\beta d(x,\tilde{x})\right) \\ p_{t+1}(\tilde{x}) = \sum_x p(x) p_t(\tilde{x}|x) \\ p_{t+1}(y|\tilde{x}) = \sum_y p(y|x) p_t(x|\tilde{x}) \end{cases} \tag{24}$$

konvergiert zu einem Minimum von $\mathcal{F}$;
Eindeutigkeit ist nicht garantiert!

# Anwendungen

- document clustering
- neural codes
- vgl. *Schneidman et. al. 2001*