The Information Bottleneck Method

T-122.102 Co-occurence methods in analysis of discrete data

Presented by Yang, Zhi-rong on 2, March 2004

Review of relevant quantization

- Problem to be solved: achieve optimal quantization rate within specified distoration tolerance

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

- quantization rate is indicated by mutual information

$$I(X;\tilde{X}) = \sum_{x\in X}\sum_{\tilde{x}\in\tilde{X}} p(x,\tilde{x})\log\left[\frac{p(\tilde{x}|x)}{p(\tilde{x})}\right]$$

Mutual information as quantization rate

$$2^{H(X)} / 2^{H(X|\tilde{X})} = 2^{I(X;\tilde{X})}$$

- the average cardinality of the partitioning of X is given by the ratio of the volume of X to that of the mean partition
- Notice that this quantity is different from the entropy of the codebook, $H(\tilde{X})$,and this entropy is normally not what we want to minimize.

Review of relevant quantization (cont.)

- Equation to be minimized:

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

or

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

- Equations for iteration:

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

Problem to be solved in this paper

- relevant information in a signal $x \in X$ and another signal $y \in Y$
- to find a short code for X that preserves the maximum information about Y
- i.e. to squeeze the information that X provides about Y through a 'bottleneck' formed by a limited set of codewords $\widetilde{X}$
- a generalization of rate distortion theory

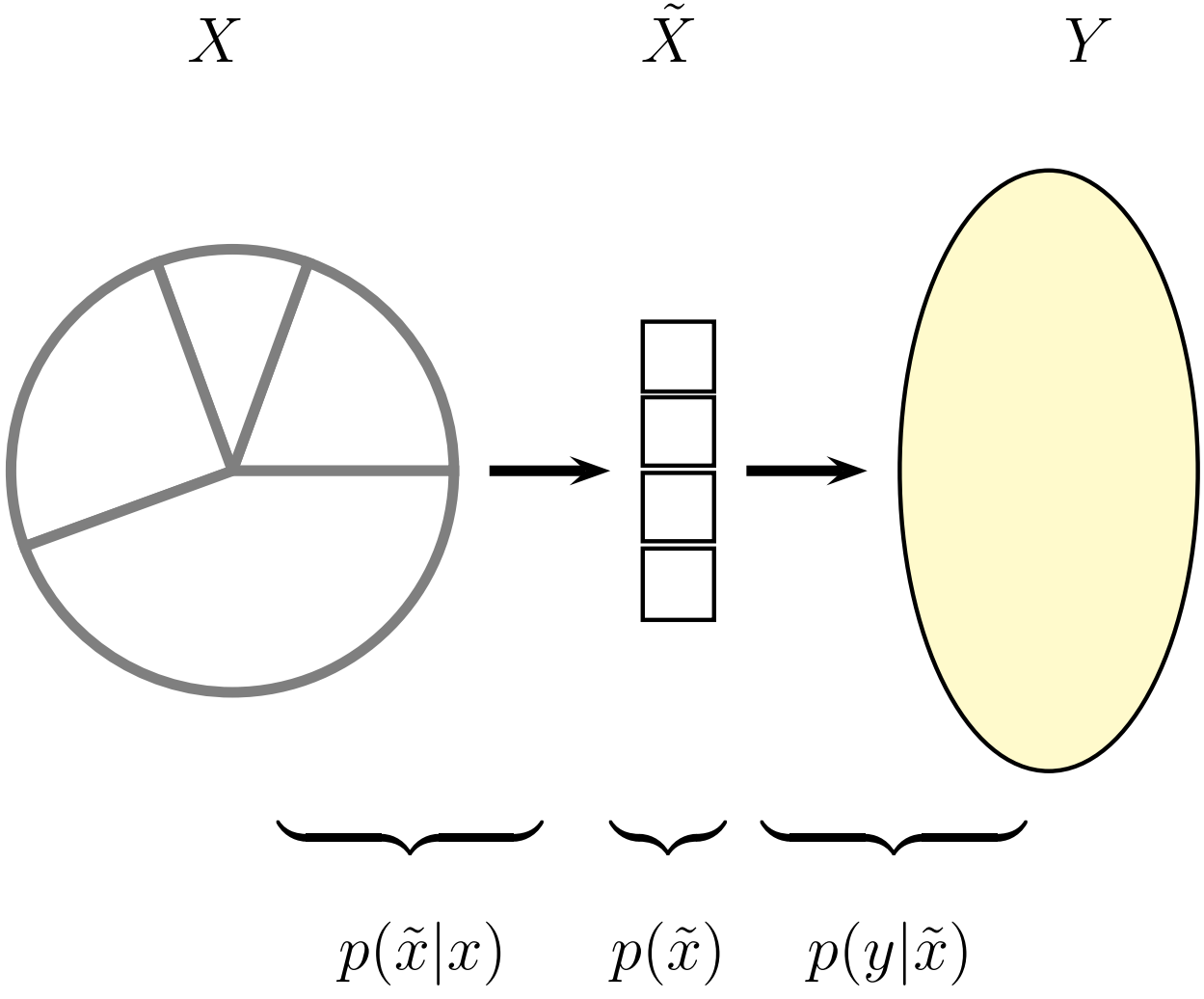
Functional to be minimized

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

where β is the Lagrange multiplier attached to the constrained meaningful information.

- $I(\tilde{X};Y) \leq I(X;Y)$ to be maximized indicates the attempt of preserving the meaningful information
- As with rate and distortion, there is a tradeoff between compressing the representation and preserving meaningful information, and there is no single right solution for the tradeoff

Passing Information through Bottleneck



Information bottleneck iterative algorithm

- Functional to be minimized:

$$\mathcal{F}\left[p(\tilde{x}|x); p(\tilde{x}); p(y|\tilde{x})\right] = I(\widetilde{X}; X) + \beta \left\langle D_{KL}\left[p(y|x)|p(y|\tilde{x})\right]\right\rangle_{p(x,\tilde{x})}$$

- Self consitent equations:

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