



HOKKAIDO UNIVERSITY

Title	連続データにも使えるMDL基準の一般化
Author(s)	鈴木, 譲
Description	ERATO 세미나2010 : No.26, 2010年11月12日
Relation	2010年度科学技術振興機構ERATO湊離散構造処理系プロジェクト講究録. p.179-185.
Issue Date	2011-06
Doc URL	https://hdl.handle.net/2115/48459
Type	conference presentation
File Information	26_all.pdf



ERATO セミナ 2010 - No. 26

連続データにも使える MDL 基準の一般化

鈴木讓

大阪大学大学院理学研究科数学専攻

2010/11/12

概要

機械学習の知識や科学の法則を発見するための基準として、MDL 原理がよく適用される。実データが与えられたときに、規則を記述し、その規則で説明しきれないデータの箇所 (例外もしくは雑音とよばれる) を記述際に、両方の記述長の合計を最小にする規則を、簡潔でデータに対して適合した規則とみなす原理である。MDL に関しては、理論的な側面は、1990 年代に (解決できるものは) ほぼ解決されたといわれている。しかしながら、ファイルやテキストのようにひずみのない (完全に復元可能な) 圧縮では、データ列の個々の記号が有限集合に含まれていることが前提となる。すなわち、MDL 原理は、実データが連続である場合には、適用されない。非常に大きな問題点でありながら、着手されず、放置されてきた問題であるといつてよい。本研究では、Radon-Nykodim 微分を用いて、記述長ではなく記述長の差を計算して、規則間の優劣を決定する方法を提案する。従来の実データが有限である場合をすべて含む一般化になっている。応用として、Bayesian ネットワークの構造学習で、一部の確率変数が連続データを含む場合も全く問題なく対応できる (自然な一般化である) ことを示す。

概要

連続データにも使える MDL 基準の一般化

鈴木 譲

大阪大学

2010 年 11 月 12 日 (金)

- ① ユニバーサルデータ圧縮
- ② 情報源の推定
- ③ ここまでのまとめ
- ④ 連続の場合のユニバーサルデータ圧縮
- ⑤ 連続の場合の情報源の推定
- ⑥ 講演のまとめ

ユニバーサルデータ圧縮

データ圧縮

A : 有限集合
 $n \in \mathbb{N} := \{1, 2, \dots\}$
 $x^n := x_1 \cdots x_n \in A^n$

データ圧縮 $c: A^n \rightarrow \{0, 1\}^*$
 $x^n \mapsto c(x^n)$

長さ $l_c: A^n \rightarrow \mathbb{N}$
 $x^n \mapsto l_c(x^n) = |c(x^n)|$ ビット数

c が一意復号可能
 c が 1 対 1

ユニバーサルデータ圧縮

Kraft の不等式

Kraft の不等式

 $l: A^n \rightarrow \mathbb{N}$

$$\sum_{x^n \in A^n} 2^{-l(x^n)} \leq 1$$

事実:

- $c: A^n \rightarrow \{0, 1\}^*$ が一意復号可能 $\implies l_c$ が Kraft の不等式
- $l: A^n \rightarrow \mathbb{N}$ が Kraft の不等式 $\implies l = l_c$ なる一意復号可能な c が存在

$Q^n(x^n) := 2^{-l(x^n)} (\iff l(x^n) = -\log Q^n(x^n))$ とおけば、

$$\sum_{x^n \in A^n} Q^n(x^n) \leq 1$$

ユニバーサルデータ圧縮

情報源

$\mathbb{Z} := \{\dots, -1, 0, 1, \dots\}$
 $\{X_i\}_{i=-\infty}^{\infty}: \dots X_1 \cdots X_n \cdots: A$ の値をとる確率変数 X_i の列

$\{X_i\}_{i=-\infty}^{\infty}$ が定常

各 $i \in \mathbb{Z}, k \in \mathbb{N}$ で

$$P(X_{i-k} \cdots X_{i-1}) = P(X_{i-k+1} \cdots X_i)$$

- X^n が i.i.d

$$P(X^n) = \prod_{i=1}^n P(X_i)$$

- X^n が k 次 Markov ($X_{-(k-1)} \cdots X_0$ に依存)

$$P(X^n) = \prod_{i=1}^n P(X_i | X_{i-k} \cdots X_{i-1})$$

ユニバーサルデータ圧縮

情報源のエントロピー

 $\{X_i\}_{i=-\infty}^{\infty}$ のエントロピー H

$$\{X_i\}_{i=-\infty}^{\infty} \text{ が定常} \implies -\frac{1}{n} \sum_{x^n \in A^n} P^n(x^n) \log P^n(x^n) \rightarrow \exists H$$

- i.i.d の場合

$$H = - \sum_{x \in A} P(x) \log P(x)$$

- k 次 Markov の場合、 $S := A^k$ として

$$H = - \sum_{s \in S} P(s) \sum_{x \in A} P(x|s) \log P(x|s)$$

$P(s)$: 状態 $s \in S$ の定常確率 (一意であることを仮定、エルゴード)

ユニバーサルデータ圧縮

データ圧縮: P が既知

- 一意復号可能な任意の c について、 $H \leq E[-l_c(X^n)]$
- $l_c(x^n) := \lceil -\log P^n(x^n) \rceil$ (切り上げ) $\implies H \leq E[l_c(X^n)] < H + 1$

ユニバーサルデータ圧縮: P が未知どのような P についても、以下が成立するような Q^n を見出す

- $\frac{1}{n} E[-\log Q^n(X^n)] \rightarrow H$
- 確率 1 で $-\frac{1}{n} \log Q^n(x^n) \rightarrow H$

後者 $\implies$ 前者 P^n が i.i.d の場合 (確率パラメータ: 未知) $m := |A|$ $c_n[x]$: $x^n \in A^n$ における $x \in A$ の頻度

$$Q^n(x^n) := \prod_{i=1}^n \frac{c_{i-1}[x_i] + \frac{1}{2}}{i - 1 + \frac{m}{2}} = \frac{\Gamma(n + \frac{m}{2}) \Gamma(\frac{1}{2})^m}{\Gamma(\frac{m}{2}) \prod_{x \in A} \Gamma(c_n[x] + \frac{1}{2})}$$

ガンマ関数

$$\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt = (x-1)\Gamma(x-1), \quad \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

$$\Gamma(n) = (n-1)!$$

 P^n が i.i.d の場合 (確率パラメータ: 未知) 続Stirling の公式: $\lim_{x \rightarrow \infty} \frac{\Gamma(x+1)}{\sqrt{2\pi x}(x/e)^x} = 1$ より、

$$L(x^n) := -\log Q^n(x^n) = n \sum_{x \in A} -\frac{c_n[x]}{n} \log \frac{c_n[x] + \frac{1}{2}}{n + \frac{m}{2}} + \frac{m-1}{2} \log n + O(1)$$

$$\sum_{x \in A} -\frac{c_n[x]}{n} \log \frac{c_n[x] + \frac{1}{2}}{n + \frac{m}{2}} \rightarrow H = \sum_{x \in A} -P(x) \log P(x) \text{ より、}$$

$$\frac{L(x^n)}{n} \rightarrow H$$

 P^n が k 次 Markov の場合 (k : 既知, 確率パラメータ: 未知) $S := A^k$ (状態集合) $c_n[x, s]$: $x^n \in A^n$ における $(x_{i-k} \cdots x_{i-1}, x_i) = (s, x) \in S \times A$ の頻度

$$Q^n(x^n|k) := \prod_{i=1}^n \frac{c_{i-1}[x_i, s_i] + \frac{1}{2}}{\{\sum_{x \in A} c_{i-1}[x, s_i]\} + \frac{m}{2}} = \prod_{s \in S} \frac{\Gamma(n + \frac{m}{2}) \Gamma(\frac{1}{2})^m}{\Gamma(\frac{m}{2}) \prod_{x \in A} \Gamma(c_n[x, s] + \frac{1}{2})}$$

同様に、 $\frac{L(x^n)}{n} = -\frac{1}{n} \log Q^n(x^n|k) \rightarrow H$ が示せる。Markov の次数 k の推定

問題

実際に生じた $x^n \in A^n$ から、それを生じた Markov の次数の k を推定 $\pi(k)$, $k = 0, 1, \dots$: Markov の次数 k の事前確率

記述長最小 = 事後確率最大

 $-\log \pi(k) - \log Q^n(x^n|k) \rightarrow \text{最小} \iff \pi(k) Q^n(x^n|k) \rightarrow \text{最大}$ Markov の次数 k が未知の場合のユニバーサルデータ圧縮

ユニバーサルデータ圧縮

$$L(x^n) := -\log \sum_{k=0}^{\infty} \pi(k) Q^n(x^n|k)$$

真の k^* が何であっても、

$$\frac{L(x^n)}{n} \leq \frac{-\log \{\pi(k^*) Q^n(x^n|k^*)\}}{n} \leq \frac{-\log \pi(k^*)}{n} + \frac{-\log Q^n(x^n|k^*)}{n} \rightarrow H$$

定常エルゴード (エルゴードな Markov)

- Bayes 符号 (今回) $\frac{L(x^n)}{n} \rightarrow H$

- Lempel-Ziv 符号 (テキスト圧縮など) $\frac{L(x^n)}{n} \rightarrow H$

同値類による条件付確率の表現

X, Y : 確率変数

$X(\Omega), Y(\Omega)$: X, Y のとりうる値 ($|Y(\Omega)| < \infty$)

問題

独立に生じた $\{(x_i, y_i)\}_{i=1}^n \in \{X(\Omega) \times Y(\Omega)\}^n$ から、 $P(Y|X)$ を推定

 $X(\Omega)$ の同値関係

$x \sim x' \iff P(y|x) = P(y|x'), y \in Y(\Omega)$

$[x]$: $x \in X(\Omega)$ の同値類

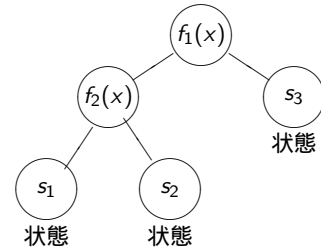
$S := \{[x] | x \in X(\Omega)\}$ (同値関係)

同値関係による条件付確率の表現: 決定木

内点 m 個、端点 l 個

$x \in X(\Omega) \rightarrow (f_1(x), \dots, f_m(x)) \rightarrow x \in \exists s \in S, |S| = l$
 $Y(\Omega) = \{0, 1\}$

$$\begin{cases} P(Y = 1 | X \in s_i) = p_i \\ P(Y = 0 | X \in s_i) = 1 - p_i \end{cases}$$



同値関係の同定

方針

実際に生じた $\{(x_i, y_i)\}_{i=1}^n \in X(\Omega) \times Y(\Omega)$ から

$|S| < \infty$ を仮定して、 $s \in S$ ごとに、 $y \in Y(\Omega)$ の頻度を計算

$c_n[y, s]$: $\{(x_i, y_i)\}_{i=1}^n$ における $y \in Y(\Omega), s \in S$ の頻度

$$Q^n(y^n | x^n, S) := \prod_{s \in S} \frac{\Gamma(n + \frac{m}{2}) \Gamma(\frac{1}{2})^m}{\Gamma(\frac{m}{2}) \prod_{y \in Y(\Omega)} \Gamma(c_n[y, s] + \frac{1}{2})}$$

$$\pi(S) Q^n(y^n | x^n, S) \rightarrow \text{最大}$$

$$-\log \pi(S) - \log Q^n(y^n | x^n, S) \rightarrow \text{最小}$$

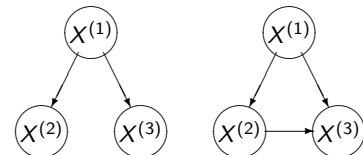
となる S を選択

Bayesian ネットワーク

$X^{(1)} < X^{(2)} < \dots < X^{(N-1)} < X^{(N)}$ (順序関係)

$$\begin{aligned} P(X^{(1)}, \dots, X^{(N)}) &= \prod_{j=1}^N P(X^{(j)} | X^{(1)}, \dots, X^{(j-1)}) \\ &= \prod_{j=1}^N P(X^{(j)} | X^{(k)}, k \in \pi^{(j)}) \end{aligned}$$

$\pi^{(j)} \subseteq \{1, \dots, j-1\}$ ($X^{(j)}$ の親集合), $j = 1, \dots, N$



$$\pi = (\{\}, \{1\}, \{1\}) \quad \pi = (\{\}, \{1\}, \{1, 2\})$$

親集合の同定

$j = 1, \dots, N$

実際に生じた $\{(x_i^{(1)}, \dots, x_i^{(j)})\}_{i=1}^n$ から、 $\pi^{(j)}$ を見出す。

- $Y = X^{(j)}$
- $(x^{(k)})_{k \in \pi^{(j)}} = s \in S := \prod_{k \in \pi^{(j)}} X^{(k)}(\Omega)$

とにおいて、 $c_n[y, s], y \in Y(\Omega), s \in S$ を計算して、

$$Q^n(x^{(j)} | \{x^{(k)}\}_{k \in \pi^{(j)}}, S) := \prod_{s \in S} \frac{\Gamma(n + \frac{m}{2}) \Gamma(\frac{1}{2})^m}{\Gamma(\frac{m}{2}) \prod_{y \in Y(\Omega)} \Gamma(c_n[y, s] + \frac{1}{2})}$$

に代入して、

$$\pi(S) Q^n(x^{(j)} | \{x^{(k)}\}_{k \in \pi^{(j)}}, S) \rightarrow \text{最大}$$

$$-\log \pi(S) - \log Q^n(x^{(j)} | \{x^{(k)}\}_{k \in \pi^{(j)}}, S) \rightarrow \text{最小}$$

なる S を選択

ここまでのまとめ

- ① データ圧縮:

$$\sum_{x^n \in A^n} Q(x^n) \leq 1$$

- ユニバーサル性: どのような P^n についても (定常エルゴード)

$$-\frac{1}{n} \log Q^n(x^n) \rightarrow H$$

を満足する Q^n を見出す (ユニバーサルデータ圧縮)

- ② $\pi(S) Q^n(x^n | k) \rightarrow \text{最大}$

$$\pi(S) Q^n(x^{(j)} | \{x^{(k)}\}_{k \in \pi^{(j)}}, S) \rightarrow \text{最大}$$

から、 k, S を推定する (情報源の推定)

連続の場合のデータ圧縮

情報理論

歪の無いデータ圧縮: 離散データ

歪のあるデータ圧縮: 連続データ

情報源が既知で、データ圧縮できても、一意復号できない

ユニバーサルデータ圧縮の定式化と一般化

 X^n : 定常エルゴード

Shannon-MacMillan-Breiman の定理

$$-\frac{1}{n} \log P^n(X^n) \rightarrow H$$

ユニバーサル性の別の定式化

 Q^n が、どのような P^n についても

$$\frac{1}{n} \log \frac{P^n(x^n)}{Q^n(x^n)} \rightarrow 0$$

連続の場合、どのように一般化されるのか

確率密度関数が存在する場合

例: $f^n :: X^n$ の確率密度関数, $X_i(\Omega) = [0, 1)$ レベル 0: $A_0 = \{[0, 1/2), [1/2, 1)\}$: 2 個レベル 1: $A_1 = \{[0, 1/4), [1/4, 1/2), [1/2, 3/4), [3/4, 1)\}$: 4 個

...

レベル k : $A_k = \{[0, 1/2^k), [1/2^k, 2/2^k), \dots, [(2^k - 1)/2^k, 1)\}$: 2^{k+1} 個

...

 $s_k: [0, 1)^n \rightarrow A_k^n$ (射影) A_k : 有限集合なので、それぞれで $Q_k^n(s_k(x^n))$ が求まる $\lambda: A_k^n \rightarrow \mathbb{R}_{\geq 0}$: $a \in A_k$ の区間の幅の $i = 1, \dots, n$ の積 (Lebesgue 測度)

$$\sum_{k=1}^{\infty} \omega_k = 1, \quad \omega_k > 0$$

$$g^n(x^n) := \sum_{k=0}^{\infty} \omega_k \frac{Q_k(s_k(x^n))}{\lambda(s_k(x^n))}$$

確率密度関数が存在する場合 続

微分エントロピー

$$h(f) := \int -f(x^n) \log f(x^n) dx^n$$

 f_k : レベル i でヒストグラム近似された分布の密度関数

Ryabko, 2009

 $\{A_k\}$ が $h(f_k^n) \rightarrow h(f^n)$ ($k \rightarrow \infty$) を満足するとき、

$$\frac{1}{n} \log \frac{f^n(x^n)}{g^n(x^n)} \rightarrow 0$$

B. Ryabko. *IEEE Trans. on Information Theory*, VOL. 55, NO. 9, 2009.

確率密度関数が存在しないとき、どうするのか

$$\int_0^\infty h(x) dx = 1$$

$$F_X(x) = \begin{cases} 0 & x < -1, \\ \frac{1}{2}, & -1 \leq x < 0 \\ \int_0^x \frac{1}{2} h(t) dt, & 0 \leq x \end{cases}$$

 $\Rightarrow F_X(x) = \int_{-\infty}^x f_X(t) dt$ なる f_X が存在しない $\mathcal{B}$: $\mathbb{R}$ における Borel 集合 X が $(\Omega, \mathcal{F}, \mu)$ における確率変数 $\mathcal{F}$ 上可測な写像 $X: \Omega \rightarrow \mathbb{R}$

$$D \in \mathcal{B} \Rightarrow \{\omega \in \Omega | X(\omega) \in D\} \in \mathcal{F}$$

Radon-Nykodim の定理

 $(\Omega, \mathcal{F}, \mu)$: 確率空間 μ は ν に関して絶対連続 ($\mu \ll \nu$)

$$\nu(A) = 0 \Rightarrow \mu(A) = 0$$

Radon-Nykodim 微分 $\frac{d\mu}{d\nu}$ $\mu \ll \nu \iff \mu(A) = \int_A g(\omega) d\nu(\omega)$ なる $\mathcal{F}$ 上可測な $g: \Omega \rightarrow \mathbb{R}$ が存在 λ : $\mathbb{R}$ における Lebesgue 測度確率変数 X に確率密度関数 f_X が存在する必要十分条件 $F_X(x) := \mu(X(\omega) \leq x)$ とおくと、

$$\mu \ll \lambda$$

Kullback-Leibler 情報量

Kullback-Leibler 情報量

 $\mu \ll \nu$ のとき

$$K(\mu||\nu) := \int \mu \log \frac{d\mu}{d\nu}$$

$$\bullet \text{ 確率 } P, Q \Rightarrow \frac{d\mu}{d\nu}(x^n) = \frac{P(x^n)}{Q(x^n)}$$

$$K(\mu||\nu) = \sum_{x^n \in A^n} P(x^n) \log \frac{P^n(x^n)}{Q^n(x^n)}$$

$$\bullet \text{ 確率密度関数 } f, g \Rightarrow \frac{d\mu}{d\nu}(x^n) = \frac{f(x^n)}{g(x^n)}$$

$$K(\mu||\nu) = \int f^n(x^n) \log \frac{f^n(x^n)}{g^n(x^n)} dx^n$$

最も一般的なユニバーサルデータ圧縮

定式化

下記の 2 条件を満足する ν^n を見出す

$$\bullet \nu^n(X^n(\Omega)) \leq 1$$

 $\bullet$ 任意の定常エルゴードな μ^n について

$$\frac{1}{n} \log \frac{d\mu^n}{d\nu^n}(x^n) \rightarrow 0$$

測度 ν^n の構成

$\{A_k\}_{k=0}^\infty$: 有限集合 A_k の系列 (A_{k+1} が A_k を分割したもの)
 $s_k: \mathbb{R} \rightarrow A_k$: A_k への射影

$Q_k^n(a^n)$, $a^n \in A_k^n$ (有限の場合のユニバーサルデータ圧縮)
 η^n : $\mu^n \ll \eta^n$ ($\eta^n = \lambda^n \Rightarrow$ Ryabko)

各 $(D_1, \dots, D_n) \in \mathcal{B}^n$ について

$$\nu_k^n(D_1, \dots, D_n) := \sum_{a_1, \dots, a_n \in A_k} \frac{\eta^n(a_1 \cap D_1, \dots, a_n \cap D_n)}{\eta^n(a_1, \dots, a_n)} Q_k^n(a_1, \dots, a_n).$$

$$\{\omega_i\}_{k=0}^\infty: \sum_{k=0}^\infty \omega_k = 1, \omega_k > 0$$

$$\nu^n(D_1, \dots, D_n) := \sum_{k=0}^\infty \omega_k \nu_k^n(D_1, \dots, D_n)$$

主定理

$$s_k(X^n) \sim P_k^n$$

$$\mu_i^n(D_1, \dots, D_n) := \sum_{a_1, \dots, a_n \in A_i} \frac{\eta^n(a_1 \cap D_1, \dots, a_n \cap D_n)}{\eta^n(a_1, \dots, a_n)} P_i^n(a_1, \dots, a_n).$$

定理

$D(\mu_i^\infty || \eta^\infty) = D(\mu^\infty || \eta^\infty)$ ($i \rightarrow \infty$) のとき、
 構成した ν^n は、任意の定常エルゴードな μ^n について

$$\frac{1}{n} \log \frac{d\mu^n}{d\nu^n}(x_1, \dots, x_n) \rightarrow 0$$

従来では実現できなかった適用例

整数のデータ圧縮

$$\Omega := \mathbb{N} = \{1, 2, \dots\}, \eta(j) = \frac{1}{j} - \frac{1}{j+1}, j \in \mathbb{N}$$

- $A_1 := \{\{1\}, \mathbb{N} - \{1\}\}$
- $A_2 := \{\{1\}, \{2\}, \mathbb{N} - \{1, 2\}\}$
- $\dots$
- $A_k := \{\{1\}, \{1, \dots, k\}, \mathbb{N} - \{1, \dots, k\}\}$
- $\dots$

 $\bullet Q_k^n(s_k(x^n))$: 有限集合 A_k の分布を推定し、

$$\frac{1}{n} \log \frac{P_k^n(s_k(x^n))}{Q_k^n(s_k(x^n))} \rightarrow 0, n \rightarrow \infty$$

 $\bullet j \in \mathbb{N} - \{1, \dots, k\}$ は、 $\eta(j) = \frac{1}{j} - \frac{1}{j+1}$ に比例して分布

一般的な条件付確率

$$(\Omega, \mathcal{F}, \mu)$$

 $A \in \mathcal{F}$ のもとでの $B \in \mathcal{F}$ の条件付確率 A によらない、 $\mathcal{F}$ 上可測な f_B が存在 (Radon-Nykodim)

$$\mu(A, B) = \int_A f(B|\omega) \mu(d\omega)$$

 $A := (X \in D), B := (Y \in D'), D, D' \in \mathcal{B}$ とおくと、

$$\mu(X \in D, Y \in D') = \int_D f(Y \in D'|x) \mu(dx)$$

 X のもとでの Y の条件付確率

$$\mu(Y \in D'|X = x) := f(Y \in D'|x) = \frac{d\mu(X \in D_x, Y \in D')}{d\mu(X \in D_x)}$$

具体的な推定方法: 同値関係の同定

$Q(y^n|x^n)$ に相当するもの

$$\nu^n(Y^n \in D^n | X^n = x^n) := \frac{d\nu(X^n \in D_{x^n}^n, Y^n \in D^n)}{d\nu(X^n \in D_{x^n}^n)}$$

$$\frac{1}{n} \log \frac{d\mu(X^n \in D_{x^n}^n, Y^n \in D^n)}{d\nu(X^n \in D_{x^n}^n, Y^n \in D^n)} \rightarrow 0$$

$$\frac{1}{n} \log \frac{d\mu(X^n \in D_{x^n}^n)}{d\nu(X^n \in D_{x^n}^n)} \rightarrow 0$$

$$\frac{1}{n} \log \frac{d\mu^n(Y^n \in D^n | X^n = x^n)}{d\nu^n(Y^n \in D^n | X^n = x^n)} = \frac{1}{n} \log \frac{\frac{d\mu(X^n \in D_{x^n}^n, Y^n \in D^n)}{d\nu(X^n \in D_{x^n}^n, Y^n \in D^n)}}{\frac{d\mu(X^n \in D_{x^n}^n)}{d\nu(X^n \in D_{x^n}^n)}} \rightarrow 0$$

親集合の同定: 一般的な場合

$j = 1, \dots, N$

実際に生じた $\{(x_i^{(1)}, \dots, x_i^{(j)})\}_{i=1}^n$ から、 $\pi^{(j)}$ を見出す。

- $Y := X^{(j)}$
- $X := \prod_{k \in \pi^{(j)}} X^{(k)}(\Omega)$
- S は、 $\pi^{(j)}$ に相当

一般的な状況

$\{X^{(j)}\}_{j=1}^N$ は、離散や連続を仮定しない

条件付確率の推定方法

アルゴリズム

同値関係 S_1, S_2 で、 $\nu^n(Y^n \in D_{y^n}^n | X^n = x^n, S_i)$, $i = 1, 2$ を計算して、

$$\frac{\pi(S_1) d\nu^n(Y^n \in D_{y^n}^n | X^n = x^n, S_1)}{\pi(S_2) d\nu^n(Y^n \in D_{y^n}^n | X^n = x^n, S_2)}$$

が 1 より大きいかな否かをみる

計算方法

- 1 (x^n, y^n) の近傍 $(D_{x^n}^n, D_{y^n}^n)$ で $\nu(X^n \in D_{x^n}^n, Y^n \in D_{y^n}^n | S)$, $\nu(X^n \in D_{x^n}^n | S)$ を計算

2

$$\nu^n(Y^n \in D_{y^n}^n | X^n = x^n, S) := \lim_{D_{x^n} \rightarrow x^n} \frac{\nu(X^n \in D_{x^n}^n, Y^n \in D_{y^n}^n | S)}{\nu(X^n \in D_{x^n}^n | S)}$$

講演のまとめ

新規な理論

- 離散や連続を仮定しないユニバーサルデータ圧縮
- 離散や連続を仮定しない Bayes/MDL 原理

応用

- Bayesian ネットワークの構造学習で、確率変数に離散と連続なものが混在
- 連続の値をとる Markov 過程の次元の同定 (X_i が $X_{i-k} \dots X_{i-1}$ のもとで X_{i-k-1} と条件付独立となる最小の k)

今後の課題

- 計算量の評価
- 単純な量子化による離散化と比較してのメリット
- 応用分野の発掘