「分類問題に対する情報理論的アプローチ」の分類整理と解説

齋藤 翔太

群馬大学

第46回情報理論とその応用シンポジウム(SITA2023)@山口県湯田温泉
2023年11月28日

私の人生初の学会（当時、修士1年生でした）

SITA 2013
ITO

第３６回 情報理論とその応用シンポジウム
2013年11月26日~29日／伊東ホテル聚楽

SITA
Since 1978

実行委員会

実行委員長	植松 友彦（東京工業大学）
プログラム委員長	松嶋 敏泰（早稲田大学）

第36回情報理論とその応用シンポジウム（SITA2013）webページより引用：https://www.ieice.org/ess/sita/SITA2013/

SITAシンポジウムの魅力

今年のSITA2013は第36回ですので、第1回は35年前の1978年の開催になります。この記念すべき第1回のSITAをご存知の先生方がSITAニューズレターにてSITAシンポジウムへの思いを語られていますので、その一部を以下に抜粋させて頂きます。ご一読頂き、泊り込みスタイルのシンポジウムの魅力の一端を感じ取って頂ければ幸です。

第36回情報理論とその応用シンポジウム（SITA2013）webページより引用：https://www.ieice.org/ess/sita/SITA2013/

SITAシンポジウムの魅力

今年のSITA2013は第36回ですので、第1回は35年前の1978年の開催になります。この記念すべき第1回のSITAをご存知の先生方がSITAニューズレターにてSITAシンポジウムへの思いを語られていますので、その一部を以下に抜粋させて頂きます。ご一読頂き、泊り込みスタイルのシンポジウムの魅力の一端を感じ取って頂ければ幸です。

【*SITA*ニューズレター*No.20*「ごあいさつ（新会長挨拶）笠原正雄（京都工芸繊維大学）より一部抜粋】

第1回は主として西の幹事がお世話をし、ホテルに泊り込んで研究討論をするという形式を選んだ。当時、電気系の研究集会活動をこのような形式で行うことは非常に珍しいことであった。何故このような泊り込み形式

第36回情報理論とその応用シンポジウム（SITA2013）webページより引用：https://www.ieice.org/ess/sita/SITA2013/

SITAシンポジウムの魅力

今年のSITA2013は第36回ですので、第1回は35年前の1978年の開催になります。この記念すべき第1回のSITAをご存知の先生方がSITAニューズレターにてSITAシンポジウムへの思いを語られていますので、その一部を以下に抜粋させて頂きます。ご一読頂き、泊り込みスタイルのシンポジウムの魅力の一端を感じ取って頂ければ幸です。

【*SITA*ニューズレター*No.20*「ごあいさつ（新会長挨拶）笠原正雄（京都工芸繊維大学）より一部抜粋】

第1回は主として西の幹事がお世話をし、ホテルに泊り込んで研究討論をするという形式を選んだ。当時、電気系の研究集会活動をこのような形式で行うことは非常に珍しいことであった。何故このような泊り込み形式

【*SITA*ニューズレター*No.23*「新会長挨拶」平澤茂一（早稲田大学）より一部抜粋】

昔の小規模だった頃には、本や論文でしか知らない偉い先生とも直接お話ができる貴重な機会がありました。また、詳細につっこんだ実のある議論も沢山あったように思います。私も第2回で発表した内容の証明法に

第36回情報理論とその応用シンポジウム（SITA2013）webページより引用：https://www.ieice.org/ess/sita/SITA2013/

【*SITA*ニューズレター*No.35*「会長あいさつ」韓太舜（電気通信大学）より一部抜粋】

「第1回情報理論とその応用研究会」は神戸の六甲荘で開かれた．参加者は総勢80名程度であったが，「その昔情報理論をやっていたことがあるので，懐かしくて参加した」という長老や大家が大半で，研究発表は30件程度に過ぎなかった．しかし，風呂に入り飯を済ませたあとは，「ワークショップ」と称して，アルコール分を十分に摂取しながら，古き良き情報理論時代の回顧やこれからの研究動向と課題などを巡って談論風発．正論，極論，激論の飛び交う中での学問的・人間的交流は誠に味わい深く，大いに意気があがり，参加者全員が古くからの知己のように親しくなってしまった．小生にとっても，このような場で得た知己が今では最も親しく最も大切な研究者仲間になっている．これは掛け替えのない生涯の財産である．

第36回情報理論とその応用シンポジウム（SITA2013）webページより引用：https://www.ieice.org/ess/sita/SITA2013/

【*SITA*ニューズレター*No.35*「会長あいさつ」韓太舜（電気通信大学）より一部抜粋】

「第1回情報理論とその応用研究会」は神戸の六甲荘で開かれた．参加者は総勢80名程度であったが，「その昔情報理論をやっていたことがあるので，懐かしくて参加した」という長老や大家が大半で，研究発表は30件程度に過ぎなかった．しかし，風呂に入り飯を済ませたあとは，「ワークショップ」と称して，アルコール分を十分に摂取しながら，古き良き情報理論時代の回顧やこれからの研究動向と課題などを巡って談論風発．正論，極論，激論の飛び交う中での学問的・人間的交流は誠に味わい深く，大いに意気があがり，参加者全員が古くからの知己のように親しくなってしまった．小生にとっても，このような場で得た知己が今では最も親しく最も大切な研究者仲間になっている．これは掛け替えのない生涯の財産である．

第36回情報理論とその応用シンポジウム（SITA2013）webページより引用：https://www.ieice.org/ess/sita/SITA2013/

SITA初参加（人生初の学会）から、ちょうど10年目に当たる
今回のSITAにおいて、基調講演という機会を頂きましたこと、
澁谷実行委員長、松本プログラム委員長をはじめとする
実行委員会、プログラム委員会の皆様に感謝申し上げます。

分類問題に対する情報理論的アプローチ

分類問題に対する情報理論的アプローチ

- カテゴリ 1に属するデータ
- カテゴリ 2に属するデータ

・・・

- カテゴリ Cに属するデータ

が与えられたもとで、
新規データ $\boldsymbol{x}$ が、どのカテゴリに属するのか
決定する問題

問題設定の詳細は次スライド

問題設定（1/3）

カテゴリ１ $\{\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}, \ldots, \boldsymbol{x}_{N_1}^{(1)}\}$

カテゴリ２ $\{\boldsymbol{x}_1^{(2)}, \boldsymbol{x}_2^{(2)}, \ldots, \boldsymbol{x}_{N_2}^{(2)}\}$

$\vdots$ $\vdots$

カテゴリ C $\{\boldsymbol{x}_1^{(C)}, \boldsymbol{x}_2^{(C)}, \ldots, \boldsymbol{x}_{N_C}^{(C)}\}$

◀ 訓練データ

新規データ $\boldsymbol{x}$ が、どのカテゴリに属するのか決定したい

問題設定（2/3）

各カテゴリに属するデータは、同一の確率分布から生起していると仮定。
ただし、**確率分布は未知**であるとする。

カテゴリ１ $\{\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}, \ldots, \boldsymbol{x}_{N_1}^{(1)}\}$ $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_{N_1}^{(1)} \sim P_1$

カテゴリ２ $\{\boldsymbol{x}_1^{(2)}, \boldsymbol{x}_2^{(2)}, \ldots, \boldsymbol{x}_{N_2}^{(2)}\}$ $\boldsymbol{x}_1^{(2)}, \ldots, \boldsymbol{x}_{N_2}^{(2)} \sim P_2$

$\vdots$ $\vdots$

カテゴリ C $\{\boldsymbol{x}_1^{(C)}, \boldsymbol{x}_2^{(C)}, \ldots, \boldsymbol{x}_{N_C}^{(C)}\}$ $\boldsymbol{x}_1^{(C)}, \ldots, \boldsymbol{x}_{N_C}^{(C)} \sim P_C$

問題設定 (3/3)

訓練データを連接した $\boldsymbol{x}^{(i)}$ $(i = 1, 2, \ldots, C)$ という表記を用いることもある

カテゴリ 1 $\{\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}, \ldots, \boldsymbol{x}_{N_1}^{(1)}\} \blacktriangleright \boldsymbol{x}^{(1)} := \boldsymbol{x}_1^{(1)} \cdots \boldsymbol{x}_{N_1}^{(1)}$

カテゴリ 2 $\{\boldsymbol{x}_1^{(2)}, \boldsymbol{x}_2^{(2)}, \ldots, \boldsymbol{x}_{N_2}^{(2)}\} \blacktriangleright \boldsymbol{x}^{(2)} := \boldsymbol{x}_1^{(2)} \cdots \boldsymbol{x}_{N_2}^{(2)}$

$\vdots$ $\vdots$

カテゴリ C $\{\boldsymbol{x}_1^{(C)}, \boldsymbol{x}_2^{(C)}, \ldots, \boldsymbol{x}_{N_C}^{(C)}\} \blacktriangleright \boldsymbol{x}^{(C)} := \boldsymbol{x}_1^{(C)} \cdots \boldsymbol{x}_{N_C}^{(C)}$

分類問題に対する情報理論的アプローチ

分類問題に対する**情報理論**的アプローチ

- 情報源符号
- 誤り訂正符号
- タイプの手法

分類問題に対する情報理論的アプローチ

A）情報源符号の考え方を利用したアプローチ
　　A-1）相対エントロピー推定に基づく方法
　　A-2）符号語長に基づく方法

B）誤り訂正符号の考え方を利用したアプローチ

C）仮説検定の形に定式化しタイプを使うアプローチ

分類問題に対する情報理論的アプローチ

A）情報源符号の考え方を利用したアプローチ

A-1）相対エントロピー推定に基づく方法

A-2）符号語長に基づく方法

B）誤り訂正符号の考え方を利用したアプローチ

C）仮説検定の形に定式化しタイプを使うアプローチ

最初に

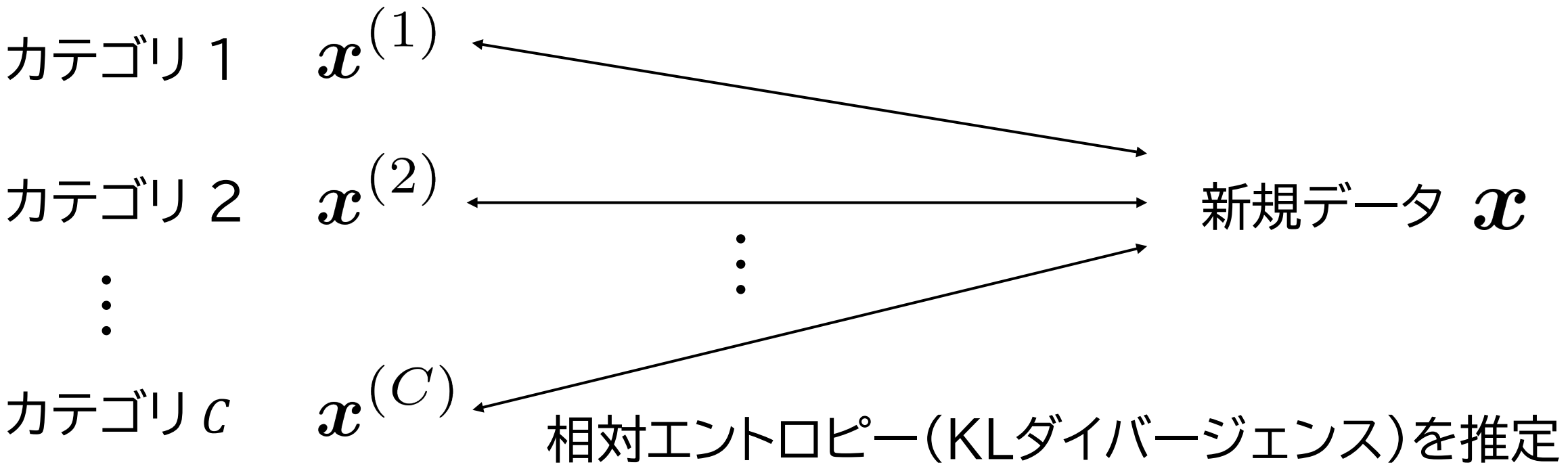
アイデア

を説明して、その後

研究例

を紹介します。

アイデア



相対エントロピーが最小となるようなカテゴリに新規データを分類

相対エントロピーを推定する際に、データ圧縮法を利用

研究例

J. Ziv and N. Merhav, ``A measure of relative entropy between individual sequences with application to universal classification,” IEEE Transactions on Information Theory, vol. 39, no. 4, pp. 1270-1279, July 1993.

考える問題

マルコフ情報源 q_z からの長さnの出力系列 $\boldsymbol{z} = z^n = z_1\, z_2 \ldots z_n$ と
マルコフ情報源 p_x からの長さnの出力系列 $\boldsymbol{x} = x^n = x_1\, x_2 \ldots x_n$
が与えられたとき、相対エントロピー

$$D(q_z \| p_x) = \sum_{s \in \mathcal{S}} q_z(s) \sum_{a \in \mathcal{A}} q_z(a|s) \log \frac{q_z(a|s)}{p_x(a|s)}$$

を推定したい。ここで、$\mathcal{S}$は状態集合、$\mathcal{A}$は情報源アルファベット。

Ziv & Merhav のアルゴリズム (1/3)

【入力】 $\boldsymbol{z} = z_1\, z_2 \ldots z_n,\quad \boldsymbol{x} = x_1\, x_2 \ldots x_n$

Ziv & Merhav のアルゴリズム (1/3)

【入力】 $\boldsymbol{z} = z_1\, z_2 \ldots z_n,\quad \boldsymbol{x} = x_1\, x_2 \ldots x_n$

【Step 1】 LZ78アルゴリズムと同様に $\boldsymbol{z}$ を**増分分解**する
（相異なるフレーズ数を $\boldsymbol{c(z)}$ と表す）

各フレーズが、それ以前に現れたフレーズと異なる最短のフレーズとなるように、系列をフレーズに分解

Ziv & Merhav のアルゴリズム (1/3)

【入力】 $\boldsymbol{z} = z_1\, z_2 \ldots z_n,\quad \boldsymbol{x} = x_1\, x_2 \ldots x_n$

【Step 1】 LZ78アルゴリズムと同様に $\boldsymbol{z}$ を**増分分解**する
（相異なるフレーズ数を $\boldsymbol{c(z)}$ と表す）

各フレーズが、それ以前に現れたフレーズと異なる最短のフレーズとなるように、系列をフレーズに分解

（例）
$\boldsymbol{z} = 01111000110$ を増分分解すると、$0, 1, 11, 10, 00, 110$ であり、$c(\boldsymbol{z}) = 6$ となる。

Ziv & Merhav のアルゴリズム (2/3)

【 Step 2】 系列 x を用いて系列 z を次のように分解する

Ziv & Merhav のアルゴリズム (2/3)

【 Step 2】 系列 $\boldsymbol{x}$ を用いて系列 $\boldsymbol{z}$ を次のように分解する

1) ある i に対して、$z_1\ z_2 \ldots z_m = x_i\ x_{i+1}\ \ldots x_{i+m-1}$ となる最大のmを見つけ、$z_1\ z_2 \ldots z_m$を系列 $\boldsymbol{z}$ の最初のフレーズとする。もし $m = 0$ ならば(すなわち、z_1が系列 $\boldsymbol{x}$ に出現しなければ)、系列 $\boldsymbol{z}$ の最初のフレーズはz_1とする。

Ziv & Merhav のアルゴリズム (2/3)

【 Step 2】 系列 $\boldsymbol{x}$ を用いて系列 $\boldsymbol{z}$ を次のように分解する

1) ある i に対して、$z_1\, z_2 \ldots z_m = x_i\, x_{i+1} \ldots x_{i+m-1}$ となる最大のmを見つけ、$z_1\, z_2 \ldots z_m$を系列 $\boldsymbol{z}$ の最初のフレーズとする。もし $m = 0$ ならば(すなわち、z_1が系列 $\boldsymbol{x}$ に出現しなければ)、系列 $\boldsymbol{z}$ の最初のフレーズはz_1とする。

2) z_{m+1}から始めて同様の手順で2番目のフレーズを見つける。

Ziv & Merhav のアルゴリズム (2/3)

【 Step 2】 系列 $\boldsymbol{x}$ を用いて系列 $\boldsymbol{z}$ を次のように分解する

1) ある i に対して、$z_1\, z_2 \ldots z_m = x_i\, x_{i+1} \ldots x_{i+m-1}$ となる最大のmを見つけ、$z_1\, z_2 \ldots z_m$を系列 $\boldsymbol{z}$ の最初のフレーズとする。もし $m = 0$ ならば(すなわち、z_1が系列 $\boldsymbol{x}$ に出現しなければ)、系列 $\boldsymbol{z}$ の最初のフレーズはz_1とする。

2) z_{m+1}から始めて同様の手順で2番目のフレーズを見つける。

3) 以下同様にして、系列 $\boldsymbol{z}$ がフレーズに分解されるまで続ける。相異なるフレーズ数を $\boldsymbol{c(z \,||\, x)}$ と表す。

Ziv & Merhav のアルゴリズム (2/3)

【 Step 2】 系列 $\boldsymbol{x}$ を用いて系列 $\boldsymbol{z}$ を次のように分解する

1) ある i に対して、$z_1\, z_2 \ldots z_m = x_i\, x_{i+1} \ldots x_{i+m-1}$ となる最大のmを見つけ、$z_1\, z_2 \ldots z_m$を系列 $\boldsymbol{z}$ の最初のフレーズとする。もし $m = 0$ ならば(すなわち、z_1が系列 $\boldsymbol{x}$ に出現しなければ)、系列 $\boldsymbol{z}$ の最初のフレーズはz_1とする。

2) z_{m+1}から始めて同様の手順で2番目のフレーズを見つける。

3) 以下同様にして、系列 $\boldsymbol{z}$ がフレーズに分解されるまで続ける。相異なるフレーズ数を $\boldsymbol{c(z \,||\, x)}$ と表す。

(例) $\boldsymbol{z} = 01111000110, \boldsymbol{x} = 10010100110$ のとき、$\boldsymbol{z}$ は $011, 110, 00110$ と分解され、$c(\boldsymbol{z}||\boldsymbol{x}) = 3$

Ziv & Merhav のアルゴリズム (3/3)

【 Step 3】 相対エントロピー $D(q_z \| p_x)$ の推定値を次式で計算

$$\frac{1}{n} c(\boldsymbol{z} \| \boldsymbol{x}) \log n - \frac{1}{n} c(\boldsymbol{z}) \log c(\boldsymbol{z})$$

Ziv & Merhav のアルゴリズム (3/3)

【 Step 3】 相対エントロピー $D(q_z \| p_x)$ の推定値を次式で計算

$$\frac{1}{n} c(\boldsymbol{z} \| \boldsymbol{x}) \log n - \frac{1}{n} c(\boldsymbol{z}) \log c(\boldsymbol{z})$$

なぜ相対エントロピーの推定値になっているかの大雑把な説明

$$\frac{1}{n} c(\boldsymbol{z} \| \boldsymbol{x}) \log n - \frac{1}{n} c(\boldsymbol{z}) \log c(\boldsymbol{z}) \approx -\frac{1}{n} \log p_x(\boldsymbol{z}) - \left(-\frac{1}{n} \log q_z(\boldsymbol{z}) \right)$$

$$= \frac{1}{n} \log \frac{q_z(\boldsymbol{z})}{p_x(\boldsymbol{z})}$$

$$\to D(q_z \| p_x) \quad (n \to \infty)$$

他の研究例

データ圧縮法を利用して、系列 $\boldsymbol{x}$ と系列 $\boldsymbol{z}$ から相対エントロピー $D(q_z \,||\, p_x)$を推定する手法としては、他にも

- **CTW法**（この後、登場します）を利用する手法
- **BWT**（Burrows-Wheeler Transform）を利用する手法

が下記の研究で提案されている。

H. Cai, S. R. Kulkarni and S. Verdu, ``Universal Divergence Estimation for Finite-Alphabet Sources,” IEEE Transactions on Information Theory, vol. 52, no. 8, pp. 3456-3475, Aug. 2006.

分類問題に対する情報理論的アプローチ

A）情報源符号の考え方を利用したアプローチ

A-1）相対エントロピー推定に基づく方法

A-2）符号語長に基づく方法

B）誤り訂正符号の考え方を利用したアプローチ

C）仮説検定の形に定式化しタイプを使うアプローチ

問題（復習）

カテゴリ１　$\boldsymbol{x}^{(1)}$

カテゴリ２　$\boldsymbol{x}^{(2)}$

⋮

カテゴリ C　$\boldsymbol{x}^{(C)}$

◀ 訓練データ

新規データ $\boldsymbol{x}$ が、どのカテゴリに属するのか決定したい

$$\boldsymbol{x}^{(i)}$$

▼ 圧縮

符号語長 $L(\boldsymbol{x}^{(i)})$

アイデア

$\boldsymbol{x}^{(i)}$

▼ 圧縮

符号語長 $L(\boldsymbol{x}^{(i)})$

連接 $\boldsymbol{x}^{(i)}\boldsymbol{x}$

▼ 圧縮

符号語長 $L(\boldsymbol{x}^{(i)}\boldsymbol{x})$

$\boldsymbol{x}^{(i)}$

▼ 圧縮

符号語長 $L(\boldsymbol{x}^{(i)})$

連接 $\boldsymbol{x}^{(i)}\boldsymbol{x}$

▼ 圧縮

符号語長 $L(\boldsymbol{x}^{(i)}\boldsymbol{x})$

差を計算 $L(\boldsymbol{x}^{(i)}\boldsymbol{x}) - L(\boldsymbol{x}^{(i)})$

▼

$i = 1, 2, \ldots, C$ に対して、この値が最小になるような
カテゴリ i に新規データ $\boldsymbol{x}$ を分類

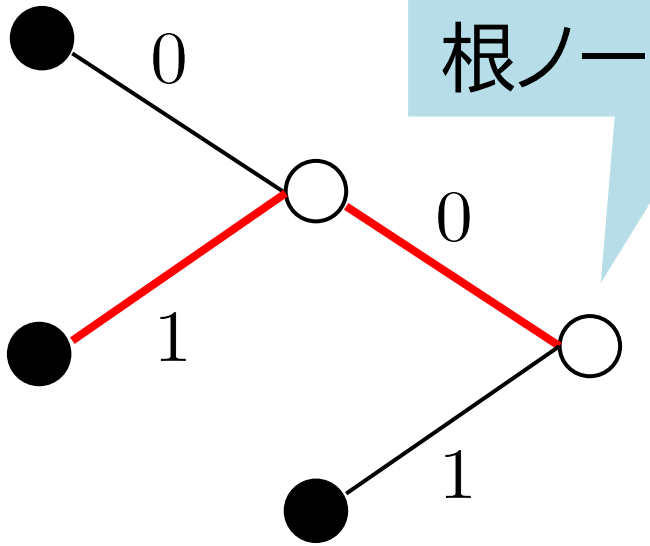
研究例

小畑智広, 池上裕之, 小林学, 坂下善彦,
``文脈木重み付け法による確率モデルを限定した文書分類,''
電子情報通信学会論文誌 D, Vol.J95-D, No.10,
pp.1873-1876, 2012.

木情報源　～状態とパラメータ～

葉ノード

根ノード



状態 s_{10}

パラメータ
$(\theta_{0|s_{10}}, \theta_{1|s_{10}})$

情報源アルファベット

$$\mathcal{A} = \{0, 1\}$$

$$\cdots \quad x_{t-2} \; x_{t-1} \; x_t$$

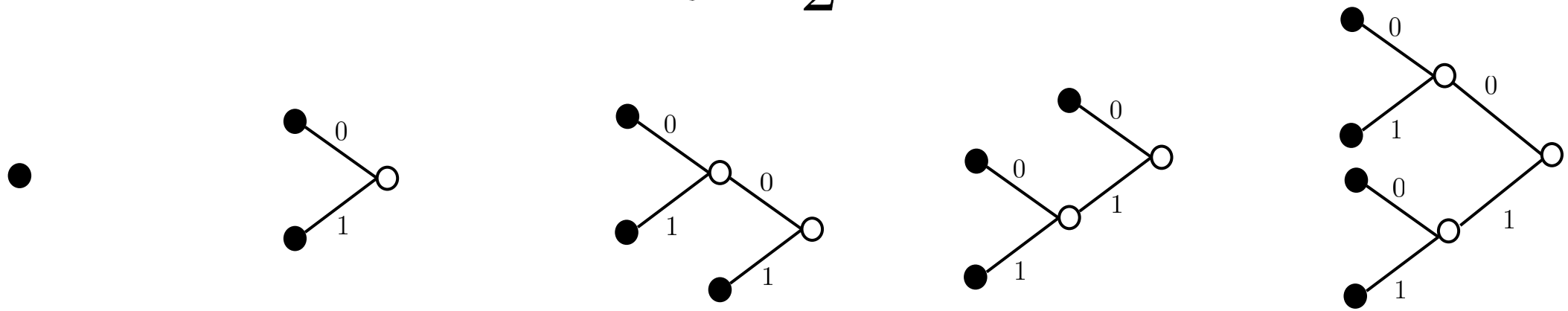
$$\cdots \quad 1 \quad 0 \quad \square$$

状態 s_{01} のもとで
- 0が生起する確率: $\theta_{0|s_{10}}$
- 1が生起する確率: $\theta_{1|s_{10}}$

木情報源　～モデルクラス～

（例）$\mathcal{A} = \{0, 1\}$，最大深さが2 → モデルクラス$\mathcal{M}_2$は以下の通り

$\mathcal{M}_2$



モデル m_1
(i.i.d.情報源)

モデル m_2
(1次のマルコフ
情報源)

モデル m_3

モデル m_4

モデル m_5
(2次のマルコフ
情報源)

最大深さがDのとき、モデルクラスを$\mathcal{M}_D$と表記

文脈木重み付け法（CTW法）(1/4)

木情報源のモデルクラス$\mathcal{M}_D$は既知であるが、真の木情報源モデルやそのパラメータは未知であると仮定。この情報源からの出力系列の圧縮法として文脈木重み付け法（CTW法）が知られている [Willems et al., 1995]

文脈木重み付け法（CTW法）(1/4)

木情報源のモデルクラス$\mathcal{M}_D$は既知であるが、真の木情報源モデルやそのパラメータは未知であると仮定。この情報源からの出力系列の圧縮法として文脈木重み付け法（CTW法）が知られている [Willems et al., 1995]

情報源アルファベット $\mathcal{A} = \{0, 1, \ldots, A-1\}$ のときのCTW法

【入力】A, D, $x_{1-D}^{0} = x_{1-D} \cdots x_0$, $x_1^n = x_1 \cdots x_n$

文脈木重み付け法（CTW法）(1/4)

木情報源のモデルクラス$\mathcal{M}_D$は既知であるが、真の木情報源モデルやそのパラメータは未知であると仮定。この情報源からの出力系列の圧縮法として文脈木重み付け法（CTW法）が知られている [Willems et al., 1995]

情報源アルファベット $\mathcal{A} = \{0, 1, \ldots, A-1\}$ のときのCTW法

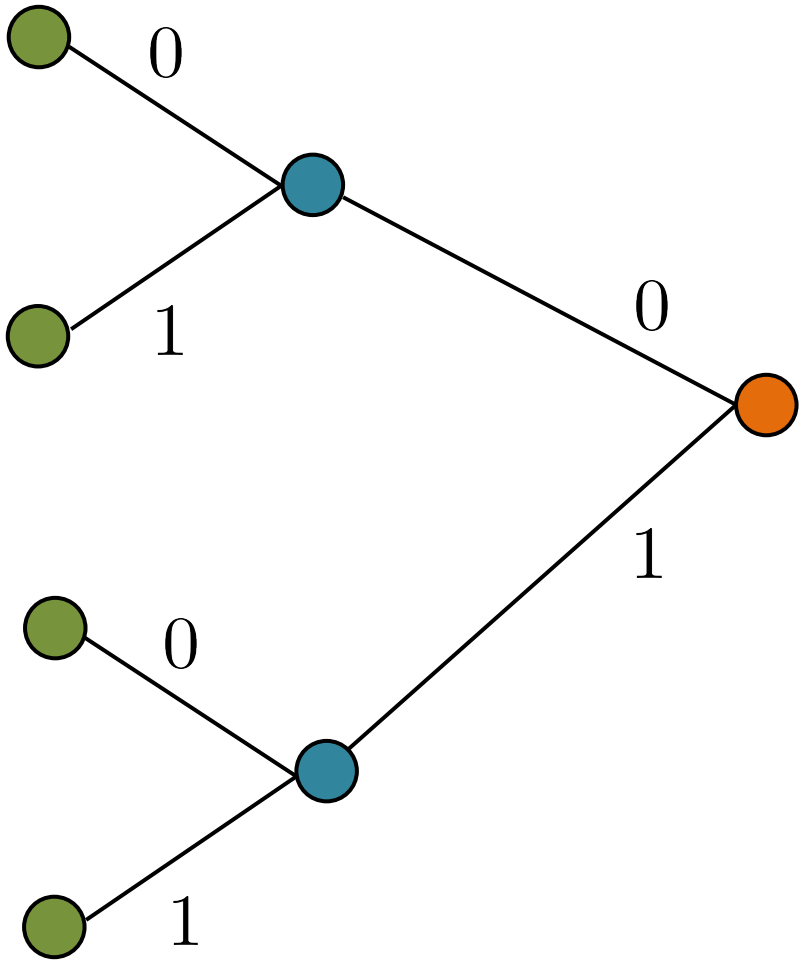
【入力】A, D, $x_{1-D}^{0} = x_{1-D} \cdots x_0$, $x_1^n = x_1 \cdots x_n$

【Step 1】各ノードsに対して $n_s = (n_s(0), n_s(1), \ldots, n_s(A-1))$ を求め（ただし、$n_s(j)$ は状態sのもとでシンボルjの生起回数）、深さ D の context tree を構成する。

（具体例は次スライド）

文脈木重み付け法（CTW法）(2/4)

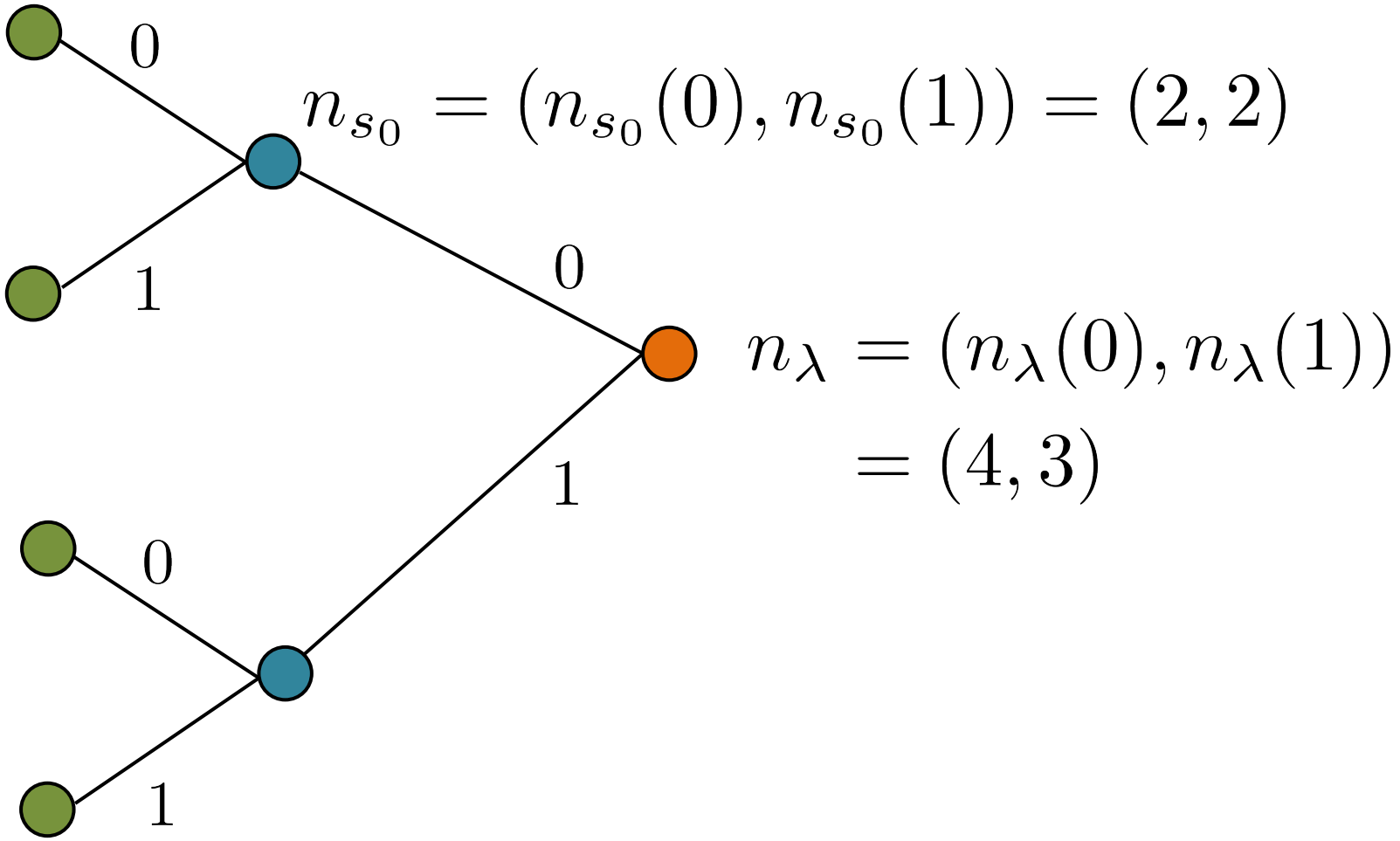
$\mathcal{A} = \{0, 1\},\ D = 2,\ x_{-1}^{0} = 10,\ x_{1}^{7} = 0110100$ のときのcontext tree



$$n_\lambda = (n_\lambda(0), n_\lambda(1)) = (4, 3)$$

文脈木重み付け法（CTW法）(2/4)

$\mathcal{A} = \{0, 1\},\ D = 2,\ x_{-1}^{0} = 10,\ x_{1}^{7} = 0110100$ のときのcontext tree



文脈木重み付け法（CTW法）(2/4)

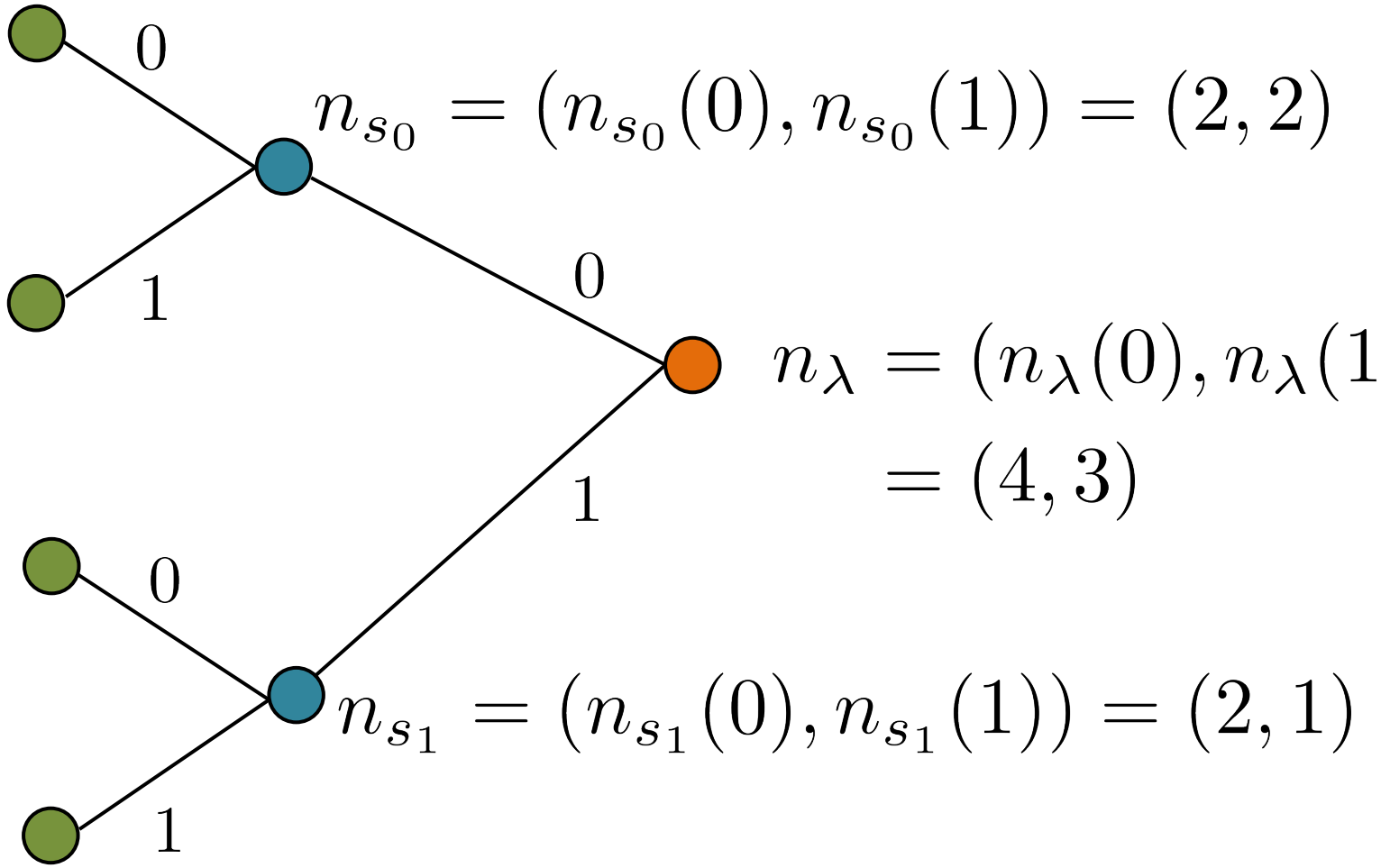
$\mathcal{A} = \{0,1\},\ D = 2,\ x_{-1}^{0} = 10,\ x_{1}^{7} = 0110100$ のときのcontext tree

$$n_{s_{00}} = (n_{s_{00}}(0), n_{s_{00}}(1)) = (0,1)$$

$$n_{s_{10}} = (n_{s_{10}}(0), n_{s_{10}}(1)) = (2,1)$$

$$n_{s_{01}} = (n_{s_{01}}(0), n_{s_{01}}(1)) = (1,1)$$

$$n_{s_{11}} = (n_{s_{11}}(0), n_{s_{11}}(1)) = (1,0)$$



$$n_{s_0} = (n_{s_0}(0), n_{s_0}(1)) = (2,2)$$

$$n_{\lambda} = (n_{\lambda}(0), n_{\lambda}(1)) = (4,3)$$

$$n_{s_1} = (n_{s_1}(0), n_{s_1}(1)) = (2,1)$$

文脈木重み付け法（CTW法）（3/4）

【Step 2】各ノードsに対して

$$P_{e,s} := \frac{\prod_{j=0}^{A-1}[(1/2)(3/2)\cdots(n_s(j)-1/2)]}{(A/2)(A/2+1)\cdots(A/2+\sum_{j=0}^{A-1} n_s(j)-1)}$$

を計算する。ただし、$n_s = (0, 0, \cdots, 0)$ のときは $P_{e,s}= 1$ とする。

例えば、$\mathcal{A} = \{0, 1\}$ (i.e., $A = 2$) のときは

$$P_{e,s} = \frac{[(1/2)(3/2)\cdots(n_s(0)-1/2)][(1/2)(3/2)\cdots(n_s(1)-1/2)]}{(n_s(0)+n_s(1))!}$$

文脈木重み付け法（CTW法）（4/4）

【Step 3】context tree の葉ノードから根ノードに向かって、
各ノードsについて次式を計算。

$$P_{w,s} := \begin{cases} P_{e,s} & (s\text{が葉ノード}) \\ \frac{1}{2}P_{e,s} + \frac{1}{2}\prod_{j=0}^{A-1} P_{w,sj} & (\text{otherwise}) \end{cases}$$

ここで、sj はsとjの連接を表す（つまりsjはノードsのj番目の子ノードに対応）

文脈木重み付け法（CTW法）（4/4）

【Step 3】context tree の葉ノードから根ノードに向かって、
各ノードsについて次式を計算。

$$P_{w,s} := \begin{cases} P_{e,s} & (s\text{が葉ノード}) \\ \frac{1}{2}P_{e,s} + \frac{1}{2}\prod_{j=0}^{A-1} P_{w,sj} & (\text{otherwise}) \end{cases}$$

ここで、sj はsとjの連接を表す（つまりsjはノードsのj番目の子ノードに対応）

【Step 4】
最後に、系列 x_{1-D}^{n}に対する符号化確率として $\boldsymbol{P_{w,\lambda}}$ を出力

ベイズ符号の視点から見ると…（1/3）

CTW法を用いた分類手法を、**ベイズ符号（Bayes code）**の視点から考察してみる。

ベイズ符号の視点から見ると…（1/3）

CTW法を用いた分類手法を、**ベイズ符号（Bayes code）**の視点から考察してみる。

- ベイズ符号・・・ベイズ決定理論に基づき、ベイズ基準のもとで最適な符号化確率を用いるユニバーサル情報源符号
  [Matsushima et al., IEEE Trans. Inf. Theory, 1991]

ベイズ符号の視点から見ると…（1/3）

CTW法を用いた分類手法を、**ベイズ符号（Bayes code）**の視点から考察してみる。

- ベイズ符号・・・ベイズ決定理論に基づき、ベイズ基準のもとで最適な符号化確率を用いるユニバーサル情報源符号
  [Matsushima et al., IEEE Trans. Inf. Theory, 1991]

- [Matsushima and Hirasawa, ISIT, 1994] において、ある形のパラメトリック事前分布を仮定することで、ベイズ符号の符号化確率を再帰計算により効率的に計算するアルゴリズムが提案されている。

ベイズ符号の視点から見ると…（1/3）

CTW法を用いた分類手法を、**ベイズ符号（Bayes code）**の視点から考察してみる。

- ベイズ符号・・・ベイズ決定理論に基づき、ベイズ基準のもとで最適な符号化確率を用いるユニバーサル情報源符号
  [Matsushima et al., IEEE Trans. Inf. Theory, 1991]

- [Matsushima and Hirasawa, ISIT, 1994] において、ある形のパラメトリック事前分布を仮定することで、ベイズ符号の符号化確率を再帰計算により効率的に計算するアルゴリズムが提案されている。

- ベイズ符号の符号語長の理論評価について、様々な研究が行われている

ベイズ符号の視点から見ると…（2/3）

既知：　木情報源のモデルクラス$\mathcal{M}_D$
未知：　真の木情報源モデル $m \in \mathcal{M}_D$、そのパラメータ θ^m
仮定：　木情報源モデルの事前分布 $P(m)$、パラメータの事前分布 $P(\theta^m | m)$

ベイズ符号の視点から見ると…（2/3）

既知：　木情報源のモデルクラス$\mathcal{M}_D$
未知：　真の木情報源モデル $m \in \mathcal{M}_D$、そのパラメータ θ^m
仮定：　木情報源モデルの事前分布 $P(m)$、パラメータの事前分布 $P(\theta^m|\, m)$

このとき、ベイズ符号の符号化確率 $P^\star(x_1^n)$ は次式で与えられる：

$$P^\star(x_1^n) = \sum_{m \in \mathcal{M}_D} P(m) \int P(x_1^n|\theta^m, m) P(\theta^m|m) d\theta^m$$

ベイズ符号の視点から見ると…（2/3）

既知：　木情報源のモデルクラス$\mathcal{M}_D$
未知：　真の木情報源モデル $m \in \mathcal{M}_D$、そのパラメータ θ^m
仮定：　木情報源モデルの事前分布 $P(m)$、パラメータの事前分布 $P(\theta^m|\, m)$

このとき、ベイズ符号の符号化確率 $P^\star(x_1^n)$ は次式で与えられる：

$$P^\star(x_1^n) = \sum_{m \in \mathcal{M}_D} P(m) \int P(x_1^n|\theta^m, m) P(\theta^m|m) d\theta^m$$

さらに、ある形の事前分布の仮定のもとで、以下が成り立つ：

$$P_{w,\lambda} = P^\star(x_1^n)$$

CTWアルゴリズムの出力

ベイズ符号の視点から見ると…（3/3）

$\boldsymbol{x}^{(i)}$ CTWで圧縮

符号語長

$$L(\boldsymbol{x}^{(i)}) = -\log P^{\star}(\boldsymbol{x}^{(i)})$$

$\boldsymbol{x}^{(i)}\boldsymbol{x}$ CTWで圧縮

符号語長

$$L(\boldsymbol{x}^{(i)}\boldsymbol{x}) = -\log P^{\star}(\boldsymbol{x}^{(i)}\boldsymbol{x})$$

ベイズ符号の視点から見ると…（3/3）

$\boldsymbol{x}^{(i)}$ CTWで圧縮

符号語長

$$L(\boldsymbol{x}^{(i)}) = -\log P^{\star}(\boldsymbol{x}^{(i)})$$

$\boldsymbol{x}^{(i)}\boldsymbol{x}$ CTWで圧縮

符号語長

$$L(\boldsymbol{x}^{(i)}\boldsymbol{x}) = -\log P^{\star}(\boldsymbol{x}^{(i)}\boldsymbol{x})$$

$$\text{分類カテゴリ} = \underset{i\in\{1,2,\ldots,C\}}{\arg\min} \left[L(\boldsymbol{x}^{(i)}\boldsymbol{x}) - L(\boldsymbol{x}^{(i)}) \right]$$

$$= \underset{i\in\{1,2,\ldots,C\}}{\arg\min} \left[-\log \frac{P^{\star}(\boldsymbol{x}^{(i)}\boldsymbol{x})}{P^{\star}(\boldsymbol{x}^{(i)})} \right] = \underset{i\in\{1,2,\ldots,C\}}{\arg\max} P^{\star}(\boldsymbol{x}|\boldsymbol{x}^{(i)})$$

ベイズ符号の視点から見ると…（3/3）

$\boldsymbol{x}^{(i)}$ CTWで圧縮

符号語長

$$L(\boldsymbol{x}^{(i)}) = -\log P^{\star}(\boldsymbol{x}^{(i)})$$

$\boldsymbol{x}^{(i)}\boldsymbol{x}$ CTWで圧縮

符号語長

$$L(\boldsymbol{x}^{(i)}\boldsymbol{x}) = -\log P^{\star}(\boldsymbol{x}^{(i)}\boldsymbol{x})$$

$$\text{分類カテゴリ} = \underset{i\in\{1,2,\ldots,C\}}{\arg\min} \left[L(\boldsymbol{x}^{(i)}\boldsymbol{x}) - L(\boldsymbol{x}^{(i)}) \right]$$

$$= \underset{i\in\{1,2,\ldots,C\}}{\arg\min} \left[-\log \frac{P^{\star}(\boldsymbol{x}^{(i)}\boldsymbol{x})}{P^{\star}(\boldsymbol{x}^{(i)})} \right] = \underset{i\in\{1,2,\ldots,C\}}{\arg\max} P^{\star}(\boldsymbol{x}|\boldsymbol{x}^{(i)})$$

訓練データ $\boldsymbol{x}^{(i)}$ が与えられた下での $\boldsymbol{x}$ の事後確率を最大とするようなクラスに分類している

分類問題に対する情報理論的アプローチ

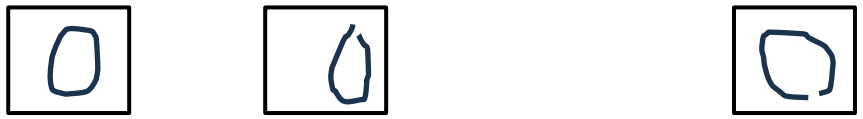
A）情報源符号の考え方を利用したアプローチ
　　A-1）符号語長に基づく方法
　　A-2）相対エントロピー推定に基づく方法

B）誤り訂正符号の考え方を利用したアプローチ

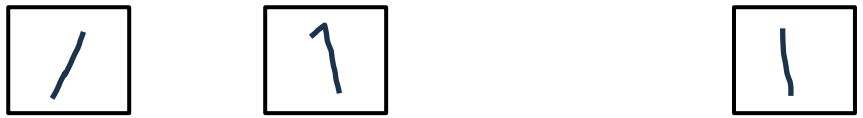
C）仮説検定の形に定式化しタイプを使うアプローチ

手書き文字認識の例

カテゴリ0 $\{\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \ldots, \boldsymbol{x}_{N_0}^{(0)}\}$

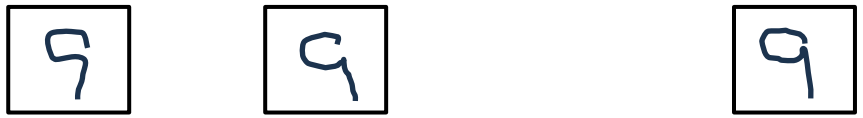


カテゴリ1 $\{\boldsymbol{x}_1^{(1)}, \boldsymbol{x}_2^{(1)}, \ldots, \boldsymbol{x}_{N_1}^{(1)}\}$



$\vdots$ $\quad$ $\vdots$

カテゴリ9 $\{\boldsymbol{x}_1^{(9)}, \boldsymbol{x}_2^{(9)}, \ldots, \boldsymbol{x}_{N_9}^{(9)}\}$



新規データ $\boldsymbol{x}$ が
どのカテゴリに
属するか決定したい

カテゴリ						
	vl	hl	dl	cc	ol	or
0	0	0	0	1	0	0
1	1	0	0	0	0	0
2	0	1	1	0	1	0
3	0	0	0	0	1	0
4	1	1	1	0	0	0
5	1	1	0	0	1	0
6	0	0	1	1	0	1
7	0	1	1	0	0	0
8	0	0	1	1	0	0
9	0	0	1	1	1	0

vl (vertical line)　hl (horizontal line)　dl (diagonal line)
cc (closed curve)　ol (curve open to left)　or (curve open to right)

カテゴリ	符号語（code word）					
	vl	hl	dl	cc	ol	or
0	0	0	0	1	0	0
1	1	0	0	0	0	0
2	0	1	1	0	1	0
3	0	0	0	0	1	0
4	1	1	1	0	0	0
5	1	1	0	0	1	0
6	0	0	1	1	0	1
7	0	1	1	0	0	0
8	0	0	1	1	0	0
9	0	0	1	1	1	0

例えば、カテゴリ0に対する符号語は 000100

カテゴリ	符号語（code word）					
	f_0	f_1	f_2	f_3	f_4	f_5
0	0	0	0	1	0	0
1	1	0	0	0	0	0
2	0	1	1	0	1	0
3	0	0	0	0	1	0
4	1	1	1	0	0	0
5	1	1	0	0	1	0
6	0	0	1	1	0	1
7	0	1	1	0	0	0
8	0	0	1	1	0	0
9	0	0	1	1	1	0

例えば、$f_0(\cdot)$は、カテゴリ1, 4, 5の$\boldsymbol{x}$については$f_0(\boldsymbol{x}) = 1$であり、
それ以外のカテゴリの$\boldsymbol{x}$については$f_0(\boldsymbol{x}) = 0$

カテゴリ	符号語（code word）					
	f_0	f_1	f_2	f_3	f_4	f_5
0	0	0	0	1	0	0
1	1	0	0	0	0	0
2	0	1	1	0	1	0
3	0	0	0	0	1	0
4	1	1	1	0	0	0
5	1	1	0	0	1	0
6	0	0	1	1	0	1
7	0	1	1	0	0	0
8	0	0	1	1	0	0
9	0	0	1	1	1	0

まず、訓練データを用いて、f_0~f_5の出力が各カテゴリの符号語で規定されるようにf_0~f_5の学習を行う。

カテゴリ	符号語（code word）					
	f_0	f_1	f_2	f_3	f_4	f_5
0	0	0	0	1	0	0
1	1	0	0	0	0	0
2	0	1	1	0	1	0
3	0	0	0	0	1	0
4	1	1	1	0	0	0
5	1	1	0	0	1	0
6	0	0	1	1	0	1
7	0	1	1	0	0	0
8	0	0	1	1	0	0
9	0	0	1	1	1	0

次に、新規データ$\boldsymbol{x}$について $f_0(\boldsymbol{x}) \sim f_5(\boldsymbol{x})$ の値を計算する。例えば $f_0(\boldsymbol{x}) = 0$, $f_1(\boldsymbol{x}) = 1, f_2(\boldsymbol{x}) = 0, f_3(\boldsymbol{x}) = 1, f_4(\boldsymbol{x}) = 0, f_5(\boldsymbol{x}) = 0$であったとすると…

カテゴリ	符号語（code word）					
	f_0	f_1	f_2	f_3	f_4	f_5
0	0	0	0	1	0	0
1	1	0	0	0	0	0
2	0	1	1	0	1	0
3	0	0	0	0	1	0
4	1	1	1	0	0	0
5	1	1	0	0	1	0
6	0	0	1	1	0	1
7	0	1	1	0	0	0
8	0	0	1	1	0	0
9	0	0	1	1	1	0

出力結果 010100 と「**最も近い**」符号語をもつカテゴリに新規データxを分類する。例としてハミング距離を考えれば、xはカテゴリ0に分類される。

カテゴリ	符号語（code word）					
	f_0	f_1	f_2	f_3	f_4	f_5
0	0	0	0	1	0	0
1	1	0	0	0	0	0
2	0	1	1	0	1	0
3	0	0	0	0	1	0
4	1	1	1	0	0	0
5	1	1	0	0	1	0
6	0	0	1	1	0	1
7	0	1	1	0	0	0
8	0	0	1	1	0	0
9	0	0	1	1	1	0

アイデア

誤り訂正符号の復号に類似→誤り訂正符号の知見を活かせば、性能の良い分類ができる？

一例として、以下の表を考えると3つまでならf_jが誤判別を起こしても誤り訂正が可能

Class	Code Word														
	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}
0	1	1	0	0	0	0	1	0	1	0	0	1	1	0	1
1	0	0	1	1	1	1	0	1	0	1	1	0	0	1	0
2	1	0	0	1	0	0	0	1	1	1	1	0	1	0	1
3	0	0	1	1	0	1	1	1	0	0	0	0	1	0	1
4	1	1	1	0	1	0	1	1	0	0	1	0	0	0	1
5	0	1	0	0	1	1	0	1	1	1	0	0	0	0	1
6	1	0	1	1	1	0	0	0	0	1	0	1	0	0	1
7	0	0	0	1	1	1	1	0	1	0	1	1	0	0	1
8	1	1	0	1	0	1	1	0	0	1	0	0	0	1	1
9	0	1	1	1	0	0	0	0	1	0	1	0	0	1	1

表は以下の文献から引用：T. G. Dietterich and G. Bakiri, ``Solving Multiclass Learning Problems via Error-Correcting Output Codes,'' Journal of Artificial Intelligence Research, Vol.2, pp.263-286, 1995.

ECOC法

誤り訂正符号に基づく多値分類法である**ECOC法**（Error-Correcting Output Coding）について、様々な研究が行われている。

ECOC法

誤り訂正符号に基づく多値分類法である**ECOC法**（Error-Correcting Output Coding）について、様々な研究が行われている。

研究例

- BCH符号等、誤り訂正符号をECOC法に援用する研究

例）T. G. Dietterich and G. Bakiri, ``Solving Multiclass Learning Problems via Error-Correcting Output Codes,'' Journal of Artificial Intelligence Research, Vol.2, pp.263-286, 1995.

ECOC法

誤り訂正符号に基づく多値分類法である**ECOC法** (Error-Correcting Output Coding)について、様々な研究が行われている。

研究例

- BCH符号等、誤り訂正符号をECOC法に援用する研究

例)T. G. Dietterich and G. Bakiri, ``Solving Multiclass Learning Problems via Error-Correcting Output Codes,'' Journal of Artificial Intelligence Research, Vol.2, pp.263-286, 1995.

- ECOC法の理論的性能評価に関する研究

例)G. Kumoi, H. Yagi, M. Kobayashi, M. Goto, and S. Hirasawa, ``Performance Evaluation of Error-Correcting Output Coding Based on Noisy and Noiseless Binary Classifiers,'' International Journal of Neural Systems, 2023.

分類問題に対する情報理論的アプローチ

A）情報源符号の考え方を利用したアプローチ
　　A-1）符号語長に基づく方法
　　A-2）相対エントロピー推定に基づく方法

B）誤り訂正符号の考え方を利用したアプローチ

C）仮説検定の形に定式化しタイプを使うアプローチ

カテゴリ 1　$\boldsymbol{x}^{(1)}$

カテゴリ 2　$\boldsymbol{x}^{(2)}$

⋮

カテゴリ C　$\boldsymbol{x}^{(C)}$

新規データ　$\boldsymbol{x}$

仮説H_1: $\boldsymbol{x}^{(1)}$ と $\boldsymbol{x}$ が同じ情報源から出力

仮説H_2: $\boldsymbol{x}^{(2)}$ と $\boldsymbol{x}$ が同じ情報源から出力

⋮

仮説H_C: $\boldsymbol{x}^{(C)}$ と $\boldsymbol{x}$ が同じ情報源から出力

訓練データ $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(C)}$ と新規データ $\boldsymbol{x}$ をもとに仮説$H_1, H_2, \ldots, H_C$ のどれを受容するか決定する問題として定式化

研究例

M. Gutman, ``Asymptotically optimal classification for multiple tests with empirically observed statistics,” IEEE Transactions on Information Theory, vol. 35, no. 2, pp. 401-408, March 1989.

考える問題

カテゴリ１　$\boldsymbol{x}^{(1)} = x_1^{(1)} x_2^{(1)} \cdots x_N^{(1)} \in \mathcal{X}^N$

i.i.d.情報源 P_1 からの長さNの出力系列

$\mathcal{X}$は
有限集合

考える問題

カテゴリ１　$\boldsymbol{x}^{(1)} = x_1^{(1)} x_2^{(1)} \cdots x_N^{(1)} \in \mathcal{X}^N$

i.i.d.情報源 P_1 からの長さNの出力系列

カテゴリ２　$\boldsymbol{x}^{(2)} = x_1^{(2)} x_2^{(2)} \cdots x_N^{(2)} \in \mathcal{X}^N$

i.i.d.情報源 P_2 からの長さNの出力系列

$\mathcal{X}$は
有限集合

考える問題

カテゴリ１　$\boldsymbol{x}^{(1)} = x_1^{(1)} x_2^{(1)} \cdots x_N^{(1)} \in \mathcal{X}^N$

i.i.d.情報源 P_1 からの長さNの出力系列

カテゴリ２　$\boldsymbol{x}^{(2)} = x_1^{(2)} x_2^{(2)} \cdots x_N^{(2)} \in \mathcal{X}^N$

i.i.d.情報源 P_2 からの長さNの出力系列

$\mathcal{X}$は有限集合

P_1, P_2 が**未知**であるとき、新規データ

$$\boldsymbol{x} = x_1 x_2 \cdots x_n \in \mathcal{X}^n$$

がカテゴリ1に属するのかカテゴリ2に属するのか決定したい。

ただし、ある $c > 0$ に対して、$N = c\,n$ とする

仮説と誤り確率

仮説H_1: $\boldsymbol{x}^{(1)}$ と $\boldsymbol{x}$ が同じ情報源から出力
仮説H_2: $\boldsymbol{x}^{(2)}$ と $\boldsymbol{x}$ が同じ情報源から出力

仮説と誤り確率

仮説H_1: $\boldsymbol{x}^{(1)}$ と $\boldsymbol{x}$ が同じ情報源から出力
仮説H_2: $\boldsymbol{x}^{(2)}$ と $\boldsymbol{x}$ が同じ情報源から出力

関数ϕ_nを $\phi_n : \mathcal{X}^N \times \mathcal{X}^N \times \mathcal{X}^n \to \{H_1, H_2\}$と定め、以下を定義:

$$\alpha(\phi_n | P_1, P_2) := \mathbb{P}_1 \left[\phi_n(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \boldsymbol{X}) = H_2 \right]$$ 第I種の誤り確率

$$\beta(\phi_n | P_1, P_2) := \mathbb{P}_2 \left[\phi_n(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \boldsymbol{X}) = H_1 \right]$$ 第II種の誤り確率

ただし、$\mathbb{P}_j[\cdot] = \Pr[\cdot \mid H_j]$

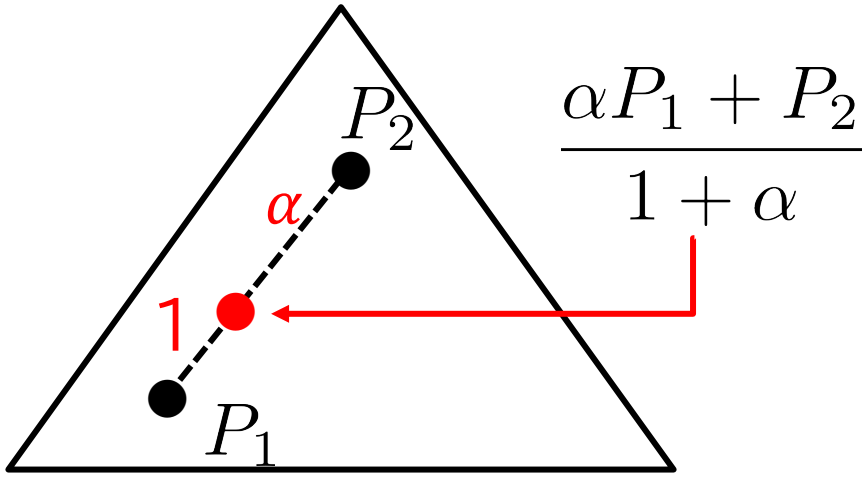
どのような関数ϕ_nを構成するか？

準備①: Generalized JS ダイバージェンス

$\alpha > 0$に対して、**generalized Jensen-Shannon ダイバージェンス**を

$$\mathrm{GJS}(P_1, P_2, \alpha) := \alpha D\left(P_1 \middle\| \frac{\alpha P_1 + P_2}{1+\alpha}\right) + D\left(P_2 \middle\| \frac{\alpha P_1 + P_2}{1+\alpha}\right)$$

と定義する。ここで、$D(\cdot \ || \cdot)$は相対エントロピー(KLダイバージェンス)



Remark　$\frac{1}{2}\mathrm{GJS}(P_1, P_2, 1)$ は、Jensen-Shannon ダイバージェンスと呼ばれる

準備②：タイプ

$\mathcal{X} = \{a_1, a_2, \ldots, a_{|\mathcal{X}|}\}, \quad \boldsymbol{x} \in \mathcal{X}^k$ とする。

系列$\boldsymbol{x}$の**タイプ**（または経験確率分布）$\hat{P}_{\boldsymbol{x}}$は、$\mathcal{X}$の各シンボルの
生起回数の相対比率である。

すなわち、

$$\hat{P}_{\boldsymbol{x}}(a_1) = \frac{N(a_1|\boldsymbol{x})}{k}, \hat{P}_{\boldsymbol{x}}(a_2) = \frac{N(a_2|\boldsymbol{x})}{k}, \cdots, \hat{P}_{\boldsymbol{x}}(a_{|\mathcal{X}|}) = \frac{N(a_{|\mathcal{X}|}|\boldsymbol{x})}{k}$$

である。ここで、$N(a\,|\boldsymbol{x})$は系列$\boldsymbol{x}$に現れるシンボル$a \in \mathcal{X}$の個数を表す。

Gutman's decision rule

$\boldsymbol{x}^{(1)} = x_1^{(1)} x_2^{(1)} \cdots x_N^{(1)}$, $\boldsymbol{x}^{(2)} = x_1^{(2)} x_2^{(2)} \cdots x_N^{(2)}$, $\boldsymbol{x} = x_1 x_2 \cdots x_n$ と

$\lambda \geq 0$が与えられたとき、[Gutman, 1989]は

$$\phi_n^{\mathrm{Gut}}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}) = \begin{cases} H_1, & h_n\left(\widehat{P}_{\boldsymbol{x}^{(1)}}, \widehat{P}_{\boldsymbol{x}}, \frac{N}{n}\right) \leq \lambda, \\ H_2, & h_n\left(\widehat{P}_{\boldsymbol{x}^{(1)}}, \widehat{P}_{\boldsymbol{x}}, \frac{N}{n}\right) > \lambda \end{cases}$$

を考えた。ここで、

$$h_n(P_1, P_2, \alpha) := \mathrm{GJS}(P_1, P_2, \alpha) + \rho(n)$$

であり、

$$\rho(n) := \frac{2|\mathcal{X}| \log(N+1) + |\mathcal{X}| \log(n+1)}{n}$$

[Gutman, 1989] の結果

定理 [Gutman, 1989]

ϕ_n は、次式を満たすとする：

$$\alpha(\phi_n|\tilde{P}_1,\tilde{P}_2) \leq 2^{-\lambda n}, \quad \forall(\tilde{P}_1,\tilde{P}_2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$$

このとき、任意の $(P_1,P_2) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ に対して、

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha(\phi_n^{\mathrm{Gut}}|P_1,P_2) \leq -\lambda$$

かつ

$$\beta(\phi_n^{\mathrm{Gut}}|P_1,P_2) \leq \beta(\phi_n|P_1,P_2)$$

他の研究例 (1/2)

下記の研究では、ϕ_n^{Gut} が「**2次オーダーの解析**」を考えたもとでも、
ある種の最適性を有していることが示されている。

L. Zhou, V. Y. F. Tan, and M. Motani,
``Second-order asymptotically optimal statistical classification,'' Information and Inference:
A Journal of the IMA, Volume 9, Issue 1,
pp. 81–111, March 2020.

他の研究例（2/2）

本日紹介した研究では、第I種の誤り確率を一定値以下としたもとで、
第II種の誤り確率について評価していた。

他の研究例（2/2）

本日紹介した研究では、第I種の誤り確率を一定値以下としたもとで、
第II種の誤り確率について評価していた。一方で、

- N. Merhav and J. Ziv, ”A Bayesian approach for classification of Markov sources,” IEEE Transactions on Information Theory, vol. 37, no. 4, pp. 1067-1071, July 1991

- S. Saito and T. Matsushima, ``Evaluation of Error Probability of Classification Based on the Analysis of the Bayes Code,'' IEEE International Symposium on Information Theory, 2020.

等では、**第I種の誤り確率と第II種の誤り確率を重み付けた「ベイズ的な誤り確率」**を分類問題の評価基準として解析を行っている。

まとめ　～分類問題に対する情報理論的アプローチ～

A）情報源符号の考え方を利用したアプローチ

A-1）相対エントロピー推定に基づく方法

A-2）符号語長に基づく方法

B）誤り訂正符号の考え方を利用したアプローチ

C）仮説検定の形に定式化しタイプを使うアプローチ

従来研究の網羅的な紹介はできませんでしたが、従来研究を整理するための
ひとつの見方を紹介しました。