

## 6. マルコフグラフ (マルコフ確率場、深層学習)

電気通信大学大学院  
情報理工学研究科  
植野 真臣

### 本日の目標

- ベイジアンネットワークの互換モデルであるマルコフネットワークについて学ぶ
- マルコフネットワークは、マルコフ確率場、条件付確率場、深層学習のボルツマン分布の上位モデルである

### マルコフネットワーク

- BNでは確率構造がDAGのときのみ、

$$P(x_1, x_2, \dots, x_N | G) = \prod_{i=1}^N p(x_i | \Pi_i, G)$$

という分解が可能になる。

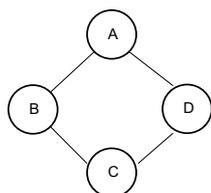
もし、確率構造がDAGにならない変数集合は分解できないので同時確率分布のまま扱わねばならない。分解できない変数集合の同時確率分布を一つのノードとしてグラフィカルモデルとして表現したものをマルコフネットワークと呼ぶ。

### 問

- $I(A, C|B, D)_G$  and  $I(B, D|A, C)_G$   
もしくは  $A \perp C|B, D$  and  $B \perp D|A, C$   
を満たすベイジアンネットワークは存在するか？

### 無向グラフでの表現

- $I(A, C|B, D)_G$  and  $I(B, D|A, C)_G$   
もしくは  $A \perp C|B, D$  and  $B \perp D|A, C$   
はベイジアンネットワークでは表現できない。



### 有向グラフから無向グラフへの変換

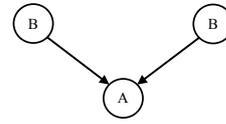
- 問
- 有向グラフの方向を取り除けば無向グラフになるのか？

## 有向グラフから無向グラフへの変換と条件付き独立性

- 問
- 有向グラフの方向を取り除けば同じ条件付き独立性を持つ無向グラフになるのか？
- NO

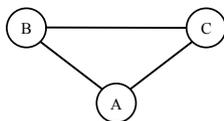
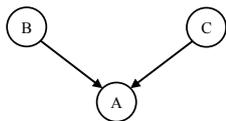
## V構造

- 無向グラフに変換せよ。



## V構造

- 無向グラフに変換せよ。

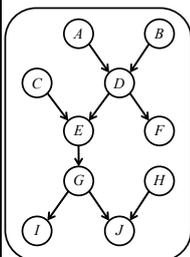


## モラルグラフ(復習)

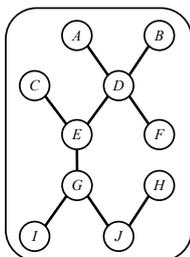
### 定義

有向グラフにおいて共通の子ノードをもつすべての親ノードの対にエッジを張り、方向性を取り除いて構築された「有向グラフに対応した無向グラフ」を**モラルグラフ**(moral graph)と呼ぶ。エッジを張ることをモラル化と呼ぶ。

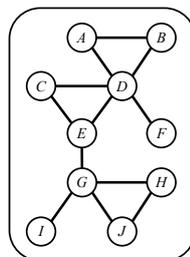
## モラルグラフの例



(a) 有向グラフ



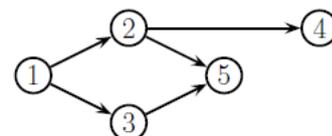
(b) 有向グラフに対応した無向グラフ



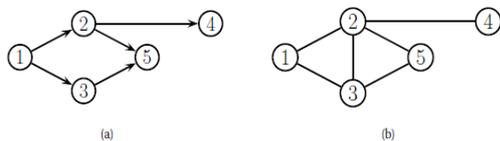
(c) 対応したモラルグラフ

モラルグラフの例

問: 以下の有向グラフを無向グラフに変換せよ。

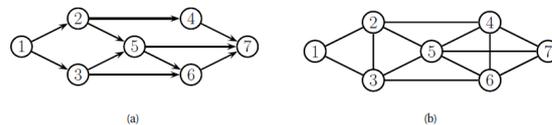


問: 以下の有向グラフを無向グラフに変換せよ。



モラル化は完全ではない

- $4 \perp 5 | 2$ を確認せよ モラル化されたグラフでは?

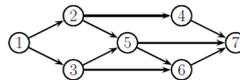


4,6にもモラル化してエッジを引いてしまったため

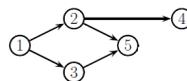
### アンセストラルグラフ(Ancestral Graph)による条件付き独立性

- 有向グラフ $G$ において $A \perp B | C$ を考える
- $A, B, C$ についてのアンセストラルグラフ(Ancestral Graph)とは $A \perp B | C$ のすべての変数集合 $U = \{A, B, C\}$ とその先祖変数集合のみによるグラフ $G$ の部分グラフのことである。言い換えれば対象となる変数とその先祖変数以外を除去する。
- $A, B, C$ についてのアンセストラルグラフを作成し、モラル化する。
- 条件付き独立性を判定する。

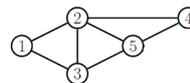
$4 \perp 5 | 2$ を確認せよ



- 2,4,5とその先祖集合 1,3を取り出す。



- モラル化



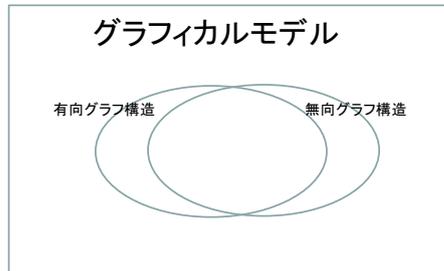
### 無向グラフでも有向グラフでも表現できる確率分布

- コーダルグラフ(Chordal Graph)によりモデル化される。
- DAGをコーダル化してクリークごとをノードに置き換えると木が作成される。DAGは必ずコーダル化できる。
- DAGが最初から木の場合、それはコーダル化しなくてもコーダルグラフである。

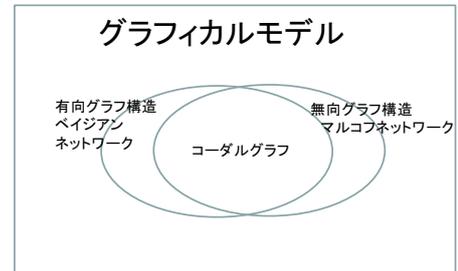
### マルコフグラフの限界

- マルコフグラフは V構造を表現できない。

## 有向グラフと無向グラフの確率分布の表現力



## 確率分布の表現



## 同時確率分布

- ベイジアンネットワークなど有向グラフではチェーンルールが適用でき、同時確率分布が表現できた。
- 無向グラフ構造でも同時確率分布が表現できるのか？

## ファクター

- 定義
- 確率変数 $x$ について、その値から実数空間への非負値への関数 $\phi$ をファクター（またはクリークポテンシャル、ポテンシャル）と呼ぶ。

## Hammersley Cliffordの定理

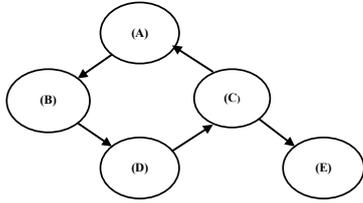
- 無向グラフ $G$ のすべての極大クリーク集合 $C$ について ファクター $\phi_c$ の積として
- $P(x_1, x_2, \dots, x_N | G) = \frac{1}{Z(\theta)} \prod_{C \in \mathcal{C}} \phi_c(x_C | \theta_C)$
- ここで $Z(\theta)$ はパーティション関数(partition function)で $x_1, x_2, \dots, x_N$ のすべての取りえる値のパターンについての和 $\sum_{x_1, x_2, \dots, x_N}$ を用いて以下のように示せる。

$$Z(\theta) \equiv \sum_{x_1, x_2, \dots, x_N} \prod_{C \in \mathcal{C}} \phi_c(x_C | \theta_C)$$

## マルコフネットワーク

- このとき
- $P(x_1, x_2, \dots, x_N | G) = \frac{1}{Z(\theta)} \prod_{C \in \mathcal{C}} \phi_c(x_C | \theta_C)$
- をギブス分布(Gibbs Distribution)と呼ぶ。
- ファクター $\phi_c$ のことを「クリークポテンシャル」、もしくは単に「ポテンシャル」ということもある。
- 無向確率ネットワークをマルコフネットワークと呼ぶ。

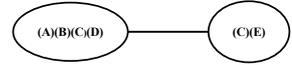
### 循環確率構造



### マルコフネットワーク

$p(A,B,C,D)$

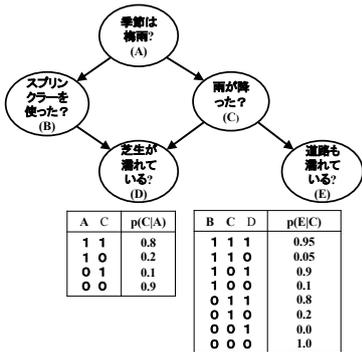
A	B	C	D	$p(A,B,C,D)$
1	1	1	1	0.06384
1	1	1	0	0.02736
1	1	0	1	0.00336
1	1	0	0	0.00144
1	1	0	1	0.02160
1	1	0	0	0.00240
1	0	1	1	0.3072
1	0	1	0	0.0768
1	0	0	1	0.0
1	0	0	0	0.09600
0	1	1	1	0.01995
0	1	1	0	0.00855
0	1	1	0	0.00105
0	1	1	0	0.00045
0	1	0	1	0.24300
0	1	0	0	0.02700
0	0	1	1	0.00800
0	0	1	0	0.00200
0	0	0	1	0.0
0	0	0	0	0.09000



$p(C,E)$

C	E	$p(C,D,E)$
1	1	0.3634
1	0	0.1560
0	1	0.0
0	0	0.4800

### ベイジアンネットワーク



A	$p(A)$
1	0.6
0	0.4

A	B	$p(B A)$
1	1	0.2
1	0	0.8
0	1	0.75
0	0	0.25

A	C	$p(C A)$
1	1	0.8
1	0	0.2
0	1	0.1
0	0	0.9

B	C	D	$p(D B,C)$
1	1	1	0.95
1	1	0	0.05
1	0	1	0.9
1	0	0	0.1
0	1	1	0.8
0	1	0	0.2
0	0	1	0.0
0	0	0	1.0

C	E	$p(E C)$
1	1	0.7
1	0	0.3
0	1	0.0
0	0	1.0

図3.12 ベイジアンネットワークのCPT<sup>1)</sup>

### マルコフネットワークの問題

- パラメータ数は指数的に増え計算量 大
- より 計算が簡易なモデルの開発
- ↓
- 隣接する変数のみを考慮するマルコフ確率場

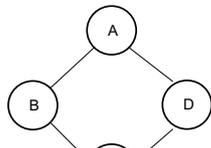
### 例題

A	B	$\phi_1(A,B)$
真	偽	1
真	真	10
偽	真	5
偽	偽	30

B	C	$\phi_2(B,C)$
真	偽	1
真	真	100
偽	真	1
偽	偽	100

C	D	$\phi_3(C,D)$
真	偽	100
真	真	1
偽	真	100
偽	偽	1

D	A	$\phi_4(D,A)$
真	偽	1
真	真	100
偽	真	1
偽	偽	100



$$A \perp C | B, D \text{ and } B \perp D | A, C$$

#### 例題

アリス(A)、ボブ(B)、チャールズ(C)、デビ(D)の4人は、話し合い授業でペアを組んで一緒に学んでいる。ある課題ごとに賛成(真)と反対(偽)の評決をとり、これまでのそれぞれの回数は上の表のとおりになった。ボブとチャールズはすごく意見が合うし、チャールズとデビはまったく意見が合わないようである。上のファクターとHammersley Cliffordの定理を用いてA, B, C, Dの同時確率分布表を作成せよ。

### 計算の仕方

$$P(A, B, C, D | G) = \frac{1}{Z} \phi_1(A, B) \cdot \phi_2(B, C) \cdot \phi_3(C, D) \cdot \phi_4(D, A)$$

ここで

$$Z = \sum_{A, B, C, D} \phi_1(A, B) \cdot \phi_2(B, C) \cdot \phi_3(C, D) \cdot \phi_4(D, A)$$

## 計算アルゴリズム

1. すべてのA,B,C,Dの取りえるパターンについて

$$\phi_1(A, B) \cdot \phi_2(B, C) \cdot \phi_3(C, D) \cdot \phi_4(D, A)$$

を計算する。

2. すべてのA,B,C,Dの取りえるパターンについて

1.で求めた結果をZとして求め、それぞれのパターンについて

$$\frac{1}{Z} \phi_1(A, B) \cdot \phi_2(B, C) \cdot \phi_3(C, D) \cdot \phi_4(D, A)$$

を求めろ。

$$\phi_1(A = 0, B = 0) \cdot \phi_2(B = 0, C = 0) \cdot \phi_3(C = 0, D = 0) \cdot \phi_4(D = 0, A = 0)$$

を求めよ。

A	B	$\phi_1(A, B)$
真偽		1
真真		10
真偽		5
偽偽		30

B	C	$\phi_2(B, C)$
真偽		1
真真		100
真偽		1
偽偽		100

C	D	$\phi_3(C, D)$
真偽		100
真真		1
真偽		100
偽偽		1

D	A	$\phi_4(D, A)$
真偽		1
真真		100
真偽		1
偽偽		100

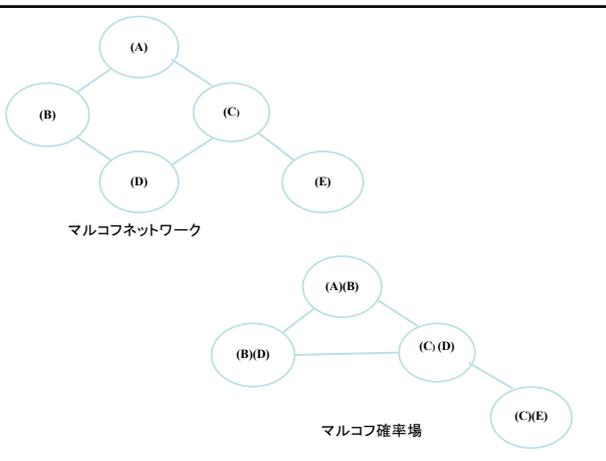
$$\phi_1(A = 0, B = 0) \cdot \phi_2(B = 0, C = 0) \cdot \phi_3(C = 0, D = 0) \cdot \phi_4(D = 0, A = 0) = 30 \cdot 100 \cdot 1 \cdot 100 = 300,000$$

## 同時確率分布表

				非正規化	正規化
A	B	C	D	$1(\cdot, \cdot) \cdot 2(\cdot, \cdot) \cdot 3(\cdot, \cdot) \cdot 4(\cdot, \cdot)$	$P(A, B, C, D)$
0	0	0	0	300,000	0.04
0	0	0	1	300,000	0.04
0	0	1	0	300,000	0.04
0	0	1	1	30	$4.1 \cdot 10^{-6}$
0	1	0	0	500	$6.9 \cdot 10^{-5}$
0	1	0	1	500	$6.9 \cdot 10^{-5}$
0	1	1	0	5,000,000	$6.9 \cdot 10^{-1}$
0	1	1	1	500	$6.9 \cdot 10^{-5}$
1	0	0	0	100	$1.4 \cdot 10^{-5}$
1	0	0	1	1,000,000	$1.4 \cdot 10^{-1}$
1	0	1	0	100	$1.4 \cdot 10^{-5}$
1	0	1	1	100	$1.4 \cdot 10^{-5}$
1	1	0	0	10	$1.4 \cdot 10^{-6}$
1	1	0	1	100,000	$1.4 \cdot 10^{-2}$
1	1	1	0	100,000	$1.4 \cdot 10^{-2}$
1	1	1	1	100,000	$1.4 \cdot 10^{-2}$
合計				7,201,840	1

## マルコフネットワークの問題

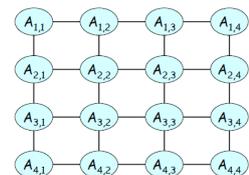
- エ夫すればベイジアンネットワークとほぼ等価に表現できる。
- しかし、パラメータ数は指数的に増え計算量大
- より 計算が簡易なモデルの開発
- ↓
- マルコフ確率場



## マルコフ確率場 (Pairwise Markov Random Field)

各ノードのファクター集合( $\phi(X_i), i = 1, \dots, N$ )と各エッジのファクター( $\phi(X_i, X_j), (X_i, X_j) \in E$ )によってファクターが定義されるモデル。クリークが2ノードのみによって構成されるという制約。

右図は完全グラフのマルコフ確率場。



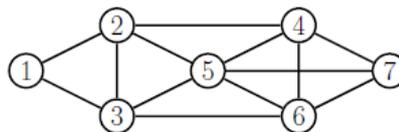
$$p(A_{1,1}, \dots, A_{4,4} | G) = \frac{1}{Z} \phi_{12}(A_{1,1}, A_{1,2}) \cdot \phi_{21}(A_{1,1}, A_{2,1}) \cdots \phi_{34}(A_{3,4}, A_{4,4})$$

## マルコフ確率場

- 確率構造に無向グラフ構造を持つグラフィカルモデルをマルコフ確率場(Markov Random Field)もしくはマルコフネットと呼ぶ。
- $I(A, B|C)_G$ : グラフGでCを除くとAとBを結ぶ路(Path)がないとき、Cを所与としてAとBは条件付き独立である。

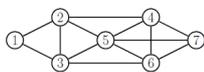
## 例

- $I(1,7|2,3)_G$  もしくは  $1 \perp 7|2,3$
- $I(1,7|4,5,6)_G$  もしくは  $1 \perp 7|4,5,6$
- $I(2,6|3,4,5)_G$  もしくは  $2 \perp 6|3,4,5$



## マルコフ ブランケット

- 定義
- ノード集合Xのマルコフ ブランケットは、Xの要素のすべての隣接ノード集合
- 例
- 5のマルコフブランケットは  $\{2,3,4,6,7\}$

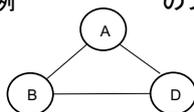


## マルコフ確率場の利点

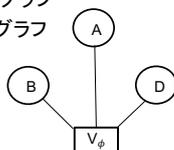
- 変数間で対象であるので、方向を持たないような変数(位置データや関係性データ)を扱う場合に自然である
- 自然言語処理における自動ラベリング問題では、隠れマルコフなどの有向グラフよりも条件付き確率場(Conditional Random Field)のほうがラベルバイアス問題がなくなり良い

## Factor Graph

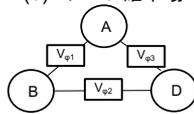
- で示される変数ノード(variable nodes)と□で示されるファクターノード(factor nodes)によって表現されるグラフで構造を可視化する。
- 例 のファクターグラフを考える。



- ファクターグラフ
- (a) マルコフグラフ



- (b) マルコフ確率場



## Factor Graphの利点

- Factor Graphはファクター構造を明示することができる。例えば、完全グラフであるN変数マルコフランダムフィールドでは、 $\binom{N}{2}$ 個のファクターノードが明示される。

## ファクターのParameterization

- ベイジアンネットワークでは、条件付き確率表がパラメータであった。ファクター $\phi_c(x_c|\theta_c)$ はどのようにパラメータ化するのであろうか？

## Log-Linear モデル

マルコフネットワークのファクターを

$$\phi_c(x_c|\theta_c) = \exp(-E(x_c|\theta_c))$$

と定義する。

ここで、 $E(x_c|\theta_c) = -\log(\phi_c(x_c|\theta_c)) > 0$ はクリーク $c$ のエネルギー関数と呼ばれる。

すなわち

$$P(x_1, x_2, \dots, x_N|G) = \frac{1}{Z(\theta)} \exp(-\sum_c E(x_c|\theta_c))$$

## マルコフ確率場のLog-Linear モデル表現

各ノードのファクター集合 $(\phi(X_i), i = 1, \dots, N)$ と各エッジのファクター $(\phi(X_i, X_j), (X_i, X_j) \in E)$ によってファクターが定義される。

$$E = \sum_i \phi(X_i) + \sum_{(i,j)} \phi(X_i, X_j)$$

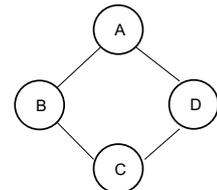
## 例題

A	B	$\phi_1(A,B)$
真偽	真偽	1
真真	真偽	10
真真	真真	5
真偽	真偽	30

B	C	$\phi_2(B,C)$
真偽	真偽	1
真真	真偽	100
真真	真真	1
真偽	真偽	100

C	D	$\phi_3(C,D)$
真偽	真偽	100
真真	真偽	1
真真	真真	100
真偽	真偽	1

D	A	$\phi_4(D,A)$
真偽	真偽	1
真真	真偽	100
真真	真真	1
真偽	真偽	100



$$A \perp C | B, D \text{ and } B \perp D | A, C$$

### 例題

アリス(A)、ポブ(B)、チャールズ(C)、デビ(D)の4人は、話し合い授業でペアを組んで一緒に学んでいる。ある課題ごとに賛成(真)と反対(偽)の評決をとり、これまでのそれぞれの回数は上の表のとおりになった。ポブとチャールズはすごく意見が合うし、チャールズとデビはまったく意見が合わないようである。上のファクターからエネルギー関数を求めよ。

例題 次のファクターから エネルギー関数を求めよ。

ファクター

A	B	$\phi_1(A,B)$
真偽	真偽	1
真真	真偽	10
真真	真真	5
真偽	真偽	30

B	C	$\phi_2(B,C)$
真偽	真偽	1
真真	真偽	100
真真	真真	1
真偽	真偽	100

A	B	$\phi_1(A,B)$
真偽	真偽	0
真真	真偽	-2.3
真真	真真	-1.61
真偽	真偽	-3.4

B	C	$\phi_2(B,C)$
真偽	真偽	0
真真	真偽	-4.61
真真	真真	0
真偽	真偽	-4.61

C	D	$\phi_3(C,D)$
真偽	真偽	100
真真	真偽	1
真真	真真	100
真偽	真偽	1

D	A	$\phi_4(D,A)$
真偽	真偽	1
真真	真偽	100
真真	真真	1
真偽	真偽	100

C	D	$\phi_3(C,D)$
真偽	真偽	-4.61
真真	真偽	0
真真	真真	-4.61
真偽	真偽	0

D	A	$\phi_4(D,A)$
真偽	真偽	0
真真	真偽	-4.61
真真	真真	0
真偽	真偽	-4.61

$$E(x_c|\theta_c) = -\log(\phi_c(x_c|\theta_c))$$

エネルギーが0は、まったく同時確率分布に寄与しないことを意味する

## Log-Linear マルコフネットワーク

一般に

- 特徴集合(a set of features)  $F = \{f_1(C_1), \dots, f_k(C_k)\}$
- $C_i$ :  $i$  番目のクリーク
- 重み集合  $w_1, \dots, w_k$

としたとき、分布 $P$ が

$$P(x_1, x_2, \dots, x_N|G) = \frac{1}{Z(\theta)} \exp\left(-\sum_{i=1}^k w_i f_i(C_i)\right)$$

で表現できるとき、Log-linear マルコフネットワークと呼ぶ。マルコフ確率場の定義を満たすものをLog-linearマルコフ確率場と呼ぶ。

## イジングモデル (Ising Model)

最初に提案されたLog-linear マルコフ確率場の一つ。原子の動作を表現するための統計物理モデル。それぞれ $\{+1, -1\}$ をとる変数集合 $(x_i)$ がある。各二変数間 $(x_i, x_j)$ のエッジのエネルギー関数は重み $w_{ij}$ として

$$E(X_i, X_j) = w_{ij} x_i x_j$$

で示される。すなわち、 $x_i = x_j$  (二つが同じ回転) のとき、 $E(X_i, X_j) = w_{ij}$  となり、それ以外のとき  $E(X_i, X_j) = -w_{ij}$  となる。

## Ising Model

一般的に以下で表現される。

$$P(x_1, x_2, \dots, x_N | G) = \frac{1}{Z(\theta)} \exp \left( - \sum_{i < j} w_{ij} x_i x_j - \sum_i u_i x_i \right)$$

$w_{ij} > 0$  のとき、モデルは  $x_i = x_j$  (二つが同じ回転) を好み、強磁性 (ferromagnetic) と呼ばれる。 $w_{ij} < 0$  のとき、モデルは  $x_i \neq x_j$  (二つが逆の回転) を好み、反強磁性 (antiferromagnetic) と呼ばれる。 $w_{ij} = 0$  のとき、二つの原子にはまったく相互作用がなく、無相互作用 (non-interaction) と呼ばれる。

## ボルツマン分布

- 熱平衡状態にある温度  $T$  の系がエネルギー  $E$  を取る確率は
- $P(E) = \frac{1}{Z} \exp(-\frac{E}{T})$
- で与えられるとき、この確率分布を Boltzmann 分布と呼ぶ。エネルギーがパラメータの線形回帰であれば指数型分布族となる。
- 当然、ギブス分布の一つである。

## ボルツマンマシン

$\{0, 1\}$  をとる変数集合  $(x_i)$  (ニューロンが発火すると 1、それ以外 0) がある。各二変数間  $(x_i, x_j)$  のエッジのエネルギー関数は重み  $w_{ij}$  として  $x_j$  の入力を所与として、以下のように発火確率が求められる。

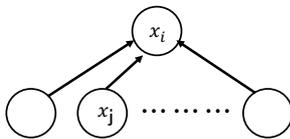
$$P(x_1, x_2, \dots, x_N | \theta) = \frac{1}{Z(\theta)} \exp \left( - \sum_{(i,j)} w_{ij} x_i x_j - \sum_i b_i x_i \right)$$

ボルツマンマシンはイジングモデルの  $\{0, 1\}$  を取るのみに修正した変形モデルである。また、隠れ変数も許す。 $(x_i, x_j)$  がともに発火した 1 の場合のみに  $\sum_{(i,j)} w_{ij} x_i x_j = 1$  となる。

## ボルツマンマシン

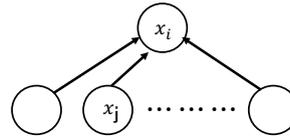
各二変数間  $(x_i, x_j)$  のエッジのエネルギー関数は重み  $w_{ij}$  として  $x_j$  の入力を所与として、以下のように発火確率が求められる。

$$P(x_i = 1 | x_j \neq x_i) = \frac{1}{1 + \exp(-\sum_j w_{ij} x_j - b_i)}$$

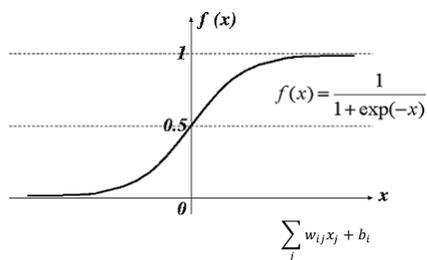


## ボルツマンマシン

$$P(x_i = 1 | x_j \neq x_i) = \frac{1}{1 + \exp(-\sum_j w_{ij} x_j - b_i)}$$

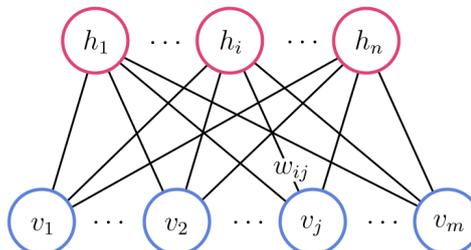


## シグモイド関数



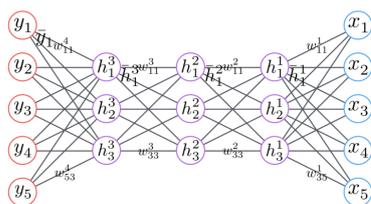
## 制限ボルツマンマシン

隠れ変数



観測変数

## 深層学習モデル

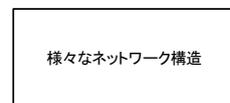


## 隠れ変数の働き

定理 Ueno (2011, UAI)

ベイジアンネットワーク学習において、疎なデータ(パラメータ数に対してデータ数が少ない)の場合、 $P(S|X)$ を最大化する構造よりも複雑な構造が必要で、最適な隠れ変数の数は増加する

○



$$P(x_1, x_2, \dots, x_N | G) = \sum_G P(G|X) P(x_1, x_2, \dots, x_N | G)$$

## 深層学習のポイント

- 隠れ変数を階層的に増加させる
- パラメータ学習を最も単純な最急降下法により、局所解への陥りを回避していること

## 隠れ変数の働き

定理 Ueno (2011, UAI)

ベイジアンネットワーク学習において、疎なデータ(パラメータ数に対してデータ数が少ない)の場合、周辺尤度最大化で得られる構造よりもより冗長な構造が真の構造である。

⇒

パターン数が多いグラフィカルモデルはほとんどがデータ数が足りない。隠れ変数を階層的に導入するとパラメータ推定が可能な範囲で爆発的にパラメータ数を増やせる。

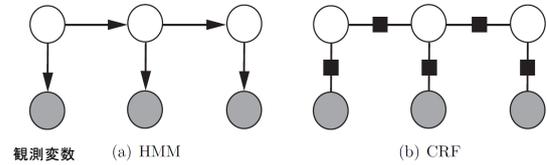
## 条件付き確率場(CRF; Conditional Random Field)

- MRFによる条件付確率の予測で分類器の一つ。

- $P(Y|X) = P(y_1, y_2, \dots, y_N | x_1, x_2, \dots, x_N)$
- $= \frac{1}{\sum_Y P(Y, X)} P(Y, X) = \frac{1}{\sum_Y P(Y, X)} \prod_i \phi_i(Y, X)$

## 隠れマルコフモデルHMM との違い

隠れ変数



観測変数 (a) HMM

(b) CRF

## HMM

- $P(X, Y) = \prod_{t=1}^T P(y_t | y_{t-1}) P(x_t | y_t)$

このモデルは同時確率分布を得ることはでき、データ発生モデルではあるが分類器(識別機)ではない。

## HEMM(McCallum et al 2000)

- HMMの識別モデル
- $P(Y|X) = \prod_{t=1}^T P(y_t | y_{t-1}, X)$   
HMMのエッジの向きを逆にしたモデル

## HMM分類器とCRF分類器

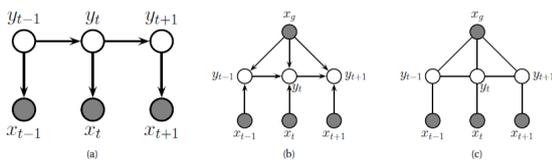
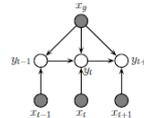


Figure 19.14 Various models for sequential data. (a) A generative directed HMM. (b) A discriminative directed MEMM. (c) A discriminative undirected CRF.

## CRF は系列ラベリングに有効

- HMMなどの有向グラフでは、後の変数データは前の変数の推論に影響しない。
- HMM分類器では  $x_t$  Biasと $y_{t-1}$ はd分離されている(Label Bias, Lafferty et al 2001)。
- HMMは文脈などが反映されない。



## 例

- 自動翻訳 (bank)
- 意味 銀行、堤防、丘

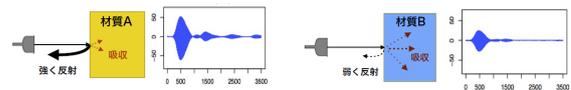
そのあとに fishという単語が出た場合、この単語がCRFでは「堤防」の意味であることがわかるが、HMMでは反映されない。

## グラフィカルモデル

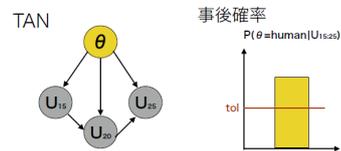
- ベイジアンネットワーク  
Naïve Bayes, Augmented Naïve Bayes, TAN, LDA, Markov model, HMM, HEMM
- マルコフネットワーク  
マルコフ確率場、(深層)ボルツマンマシン、条件付き確率場

## 応用研究

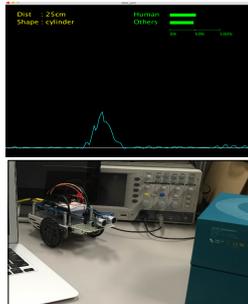
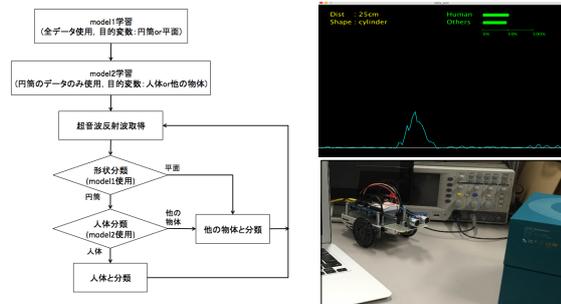
超音波データ解析に基づく  
ベイズ的人体検出ロボットの開発  
超音波の反射波の形状の違いをもとに人体を検出



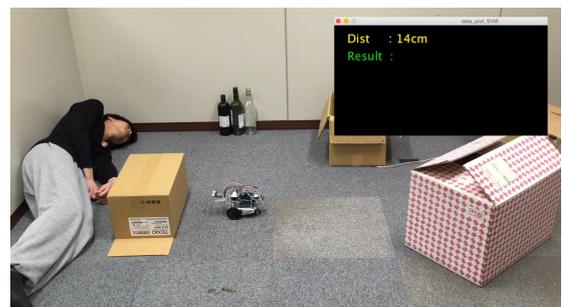
反射波のデータから学習した  
ベイジアンネットワーク分類器により判定



## 超音波データ解析に基づく ベイズ的人体検出ロボットの開発



災害時などの視界が悪い環境下で有効



## 顔表情認識技術の開発

カメラから得られる目・鼻・口などの特徴点のデータから、

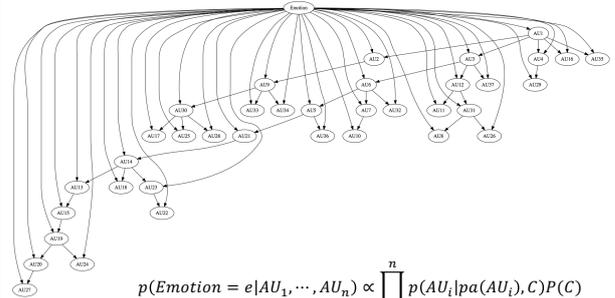


特徴量を変数とするベイジアンネットワーク分類器を学習し、

それを用いて表情判定



## 学習したベイジアンネットワーク分類器



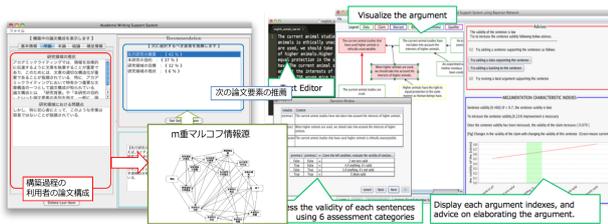
$$p(\text{Emotion} = e | AU_1, \dots, AU_n) \propto \prod_{i=1}^n p(AU_i | pa(AU_i), C) P(C)$$

ネットワーク構造を所与とした

目的変数の事後確率から表情を推定

89%の正確度

## 自然言語処理



自然言語処理、線形計画法とベイジアンネットワークにより文章の論理構造を自動理解するシステムの開発  
論述式試験の自動評価などに適用可能

## ビッグデータ時代のデータサイエンス

- データ数の増加により十分なデータが与えられていることを意味しない。
- 精緻な現象予測のために、変数の組み合わせ爆発が起こり、計算量の爆発、データ不足を如何に解決するかはの学問。

• ご清聴ありがとうございました！！