

ベイズ推論による数字パターン 認識

2020.09.04

(有)数理解析研究所 一色 浩

SAPT研究発表会

1. はじめに

ディープラーニング(DL)の登場により、人間には簡単にできるのに機械ではできなかった画像認識、音声認識、機械翻訳などの技術にブレークスルーが起きた。

DLにより、ニューラル・ネットワーク(NN)の大規模化とコンボリューションネットワークなどの新規化が可能となり、学習能力が飛躍的に向上したためであり、NNあるいはDLは人工知能(AI)の代名詞となり、遠からず人間の知的作業はすべてAIで可能となるという見方すら生まれるようになった。

しかし、DLには、「推論がブラックボックスであること」、「過学習による想定外の答え」、「ネットワークの大規模化と長時間学習」などの問題がある。

ベイズ推論は、NNとはまったく異なる学習および推論を提供でき、NNの持つ問題を産まない可能性がある。本論では、ベイズ推論に基づくパターン認識を論ずる。

2. ベイズの定理とは？

原因 θ の下での結果 D の確率: $P(D | \theta) = \frac{P(D, \theta)}{P(\theta)}$

結果 D の下での原因 θ の確率(ベイズの定理):

$$P(\theta | D) = \frac{P(D, \theta)}{P(D)} = \frac{P(D | \theta)P(\theta)}{P(D)}$$

$P(\theta)$ を事前確率、 $P(\theta | D)$ を事後確率、 $P(D | \theta)$ を尤度関数

ベイズの推論とは？

条件付き確率 $P(D | \theta)$ を既知として、事前確率 $P(\theta)$ の推定に基づいて、逆確率 $P(\theta | D)$ を推論すること:

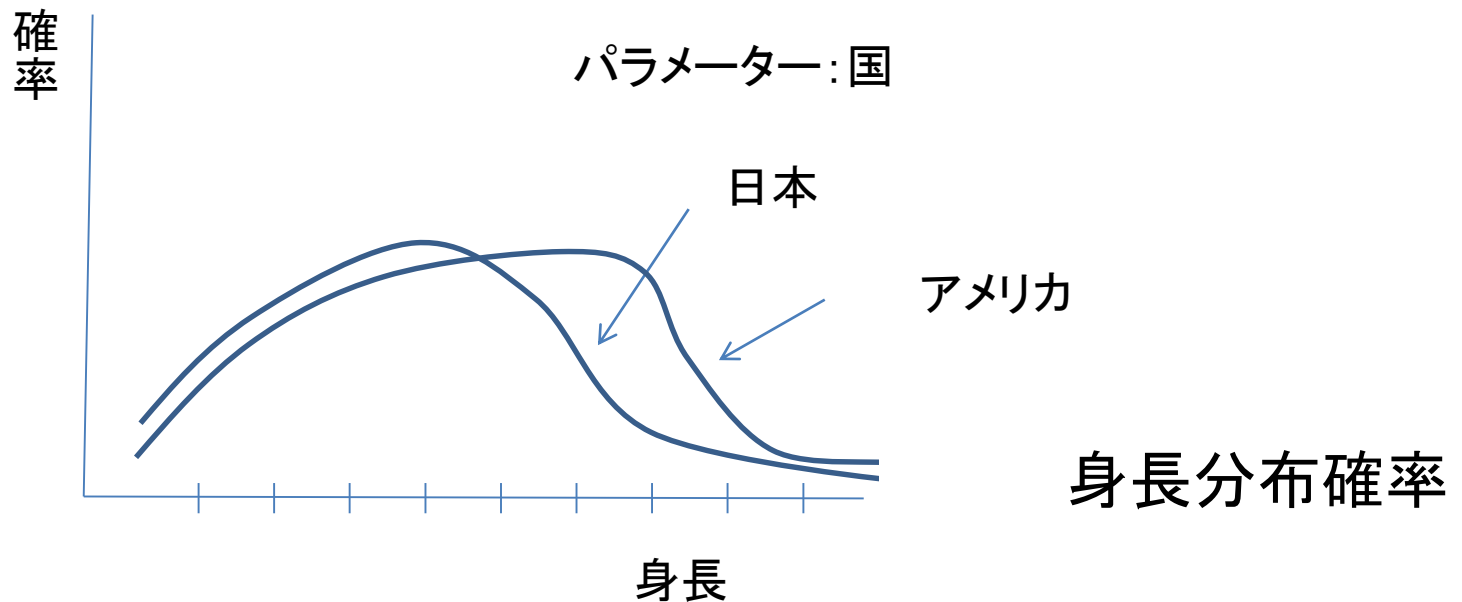
$$P(\theta | D) = \frac{P(D, \theta)}{P(D)} = \frac{P(D, \theta)}{\sum_B P(D, \theta)} \text{ where } P(D, \theta) = P(D | \theta)P(\theta)$$

ベイズの学習とは？ ... $P(D | \theta)$ を求めること

$P(A)$ に常識(主観)を使うので、一時期忌避されたが、今ではそれを特長として復活。

国籍別の身長分布確率がある。身長から国籍を推定するには？

$$P(\text{身長} \mid \text{国籍}) \rightarrow P(\text{国籍} \mid \text{身長})$$



1. 事前確率...各国人(米人...)である確率
2. データ確率... $P(\text{身長区分} \mid \text{各国人})$
3. データ確率 $\times$ 事前確率 $\rightarrow$ 同時確率 $\rightarrow P(\text{各国人} \mid \text{身長区分})$...例えば、 $P(\text{アメリカ人} \mid 180\text{cm} \sim 200\text{cm})$

国籍別の体重分布確率がある。体重から国籍を推定するには？

事前確率...gvn: $P(Nat_n)$

データ確率...gvn: $P(Wgt_i | Nat_n)$

cal...同時確率: $P(Wgt_i, Nat_n) = P(Wgt_i | Nat_n)P(Nat_n)$

cal...逆確率: $P(Nat_n | Wgt_i) = \frac{P(Wgt_i, Nat_n)}{P(Wgt_i)} = \frac{P(Wgt_i, Nat_n)}{\sum_{n=1}^N P(Wgt_i, Nat_n)}$

国籍別の身長と体重それぞれの分布確率がある。
身長と体重から国籍を推定するには？

事前確率...gvn: $P(Nat_n)$

データ確率...gvn: $P(Hgt_i, Wgt_j | Nat_n)$

cal...同時確率: $P(Hgt_i, Wgt_j, Nat_n) = P(Hgt_i, Wgt_j | Nat_n)P(Nat_n)$

cal...ベイズ推定:

$$P(Nat_n | Hgt_i, Wgt_j) = \frac{P(Hgt_i, Wgt_j, Nat_n)}{P(Hgt_i, Wgt_j)} = \frac{P(Hgt_i, Wgt_j, Nat_n)}{\sum_{n=1}^N P(Hgt_i, Wgt_j, Nat_n)}$$

国籍別の身長、体重、目の色、髪の色、肌の色それぞれの分布確率がある。身長、体重、目の色、髪の色、肌の色から国籍を推定するには？

3. 数字パターン認識への応用

手書き数字をBayes推定で認識

1. 手書き数字パターン(MNIST)を、例えば 28×28 の格子パターンで0、1データに変換する。
2. 多数の学習用手書き数字パターンを使って各格子の0,1の確率分布 $P(\text{Cell}_i | \text{Num}_n)$ を求める。
3. 判定用数字デジタルデータの条件付き同時確率分布 $P(\text{Cell}_1, \text{Cell}_2, \dots, \text{Cell}_I | \text{Num}_n)$ を求める。
4. 上の条件付き同時確率分布から、同時確率分布 $P(\text{Cell}_1, \text{Cell}_2, \dots, \text{Cell}_I, \text{Num}_n)$ を求める。
5. 判定用数字デジタルデータを条件とする手書き数字の確率分布 $P(\text{Num}_n | \text{Cell}_1, \text{Cell}_2, \dots, \text{Cell}_I)$ を求める。

事前確率...gvn: $P(Num_n)$

データ確率...gvn:

$$\begin{aligned} &P(Cell_1, Cell_2, \dots, Cell_I \mid Num_n) \\ &= P(Cell_1 \mid Num_n)P(Cell_2 \mid Num_n) \cdots P(Cell_I \mid Num_n) \end{aligned}$$

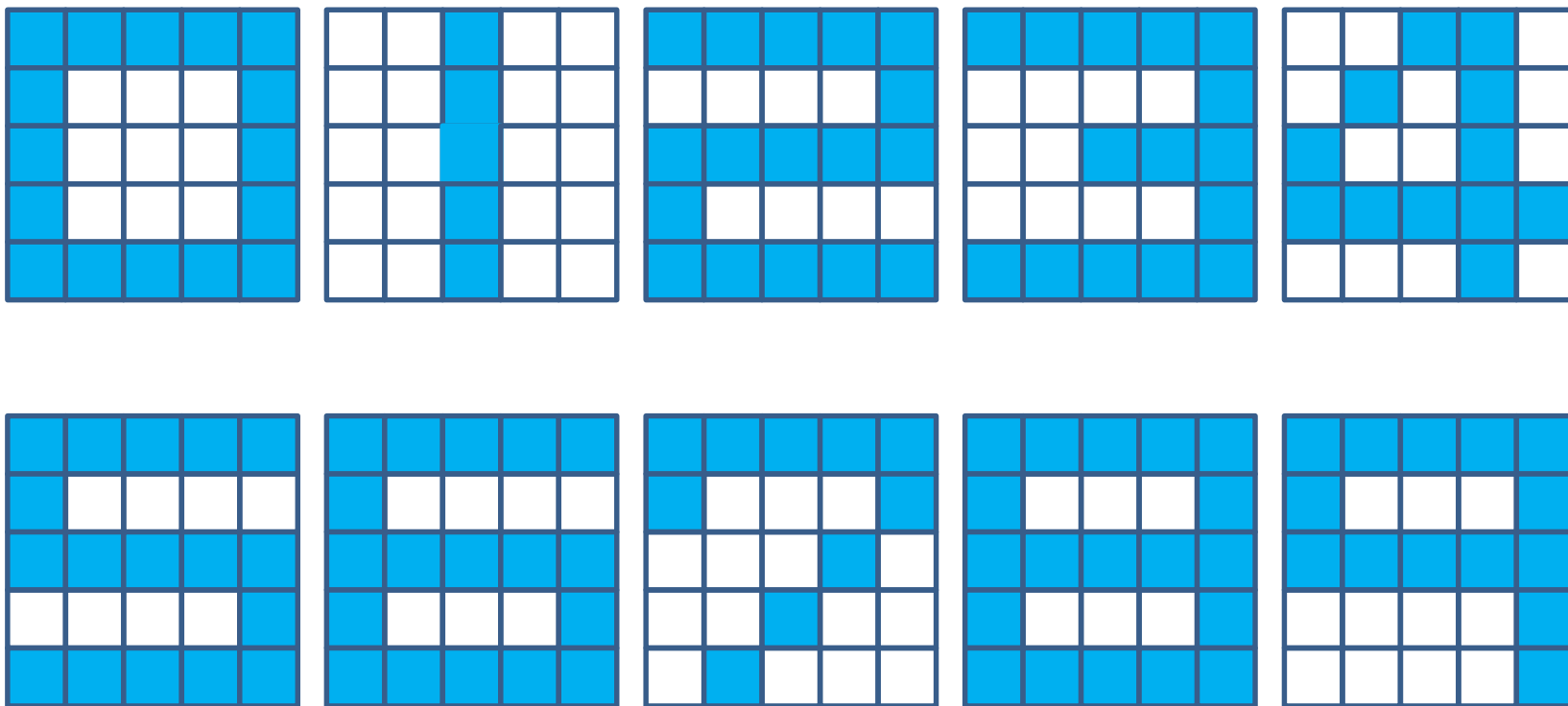
cal...同時確率 $P(Cell_1, Cell_2, \dots, Cell_I, Num_n)$

$$= P(Cell_1, Cell_2, \dots, Cell_I \mid Num_n)P(Num_n)$$

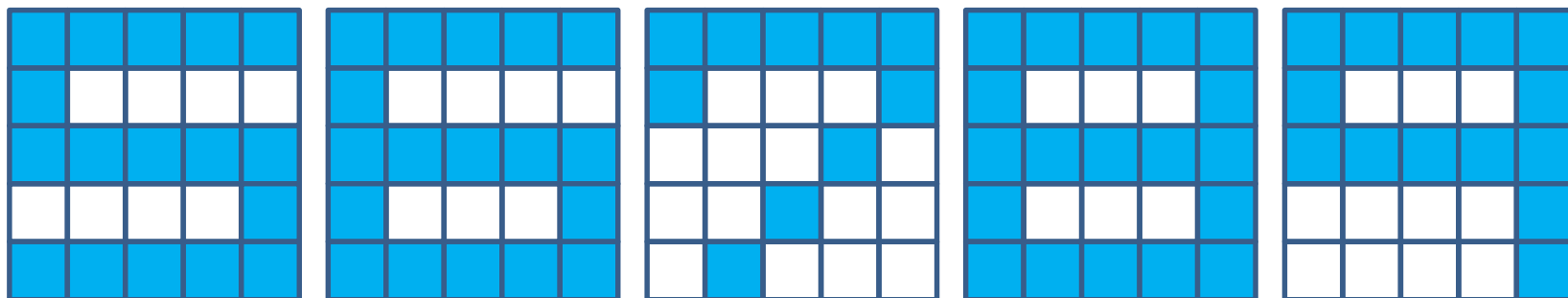
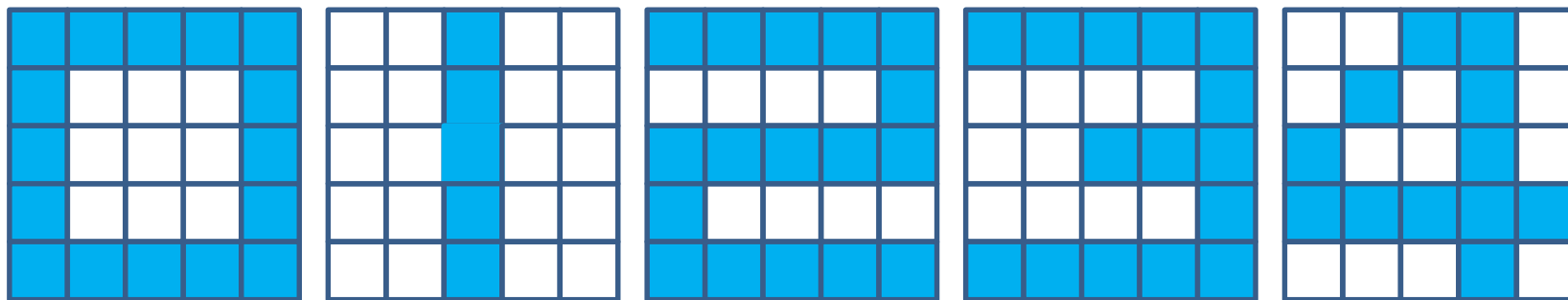
cal...逆確率:

$$\begin{aligned} P(Num_n \mid Cell_1, Cell_2, \dots, Cell_I) &= \frac{P(Cell_1, Cell_2, \dots, Cell_I, Num_n)}{P(Pgt_1, Pgt_2, \dots, Pgt_I)} \\ &= \frac{P(Cell_1, Cell_2, \dots, Cell_I, Num_n)}{\sum_{n=1}^N P(Cell_1, Cell_2, \dots, Cell_I, Num_n)} \end{aligned}$$

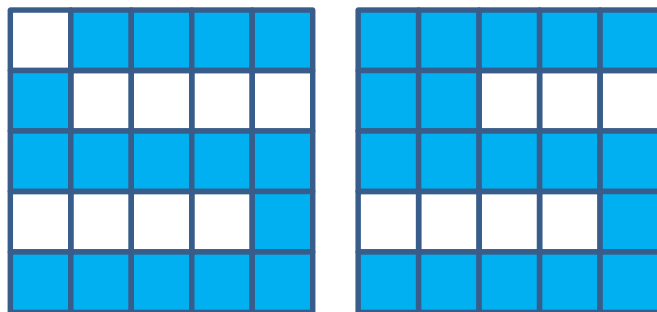
学習用データ(数字認識)



正しく認識されたデータ

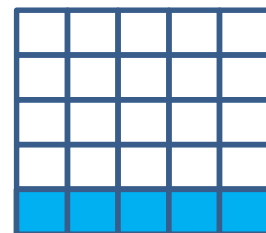
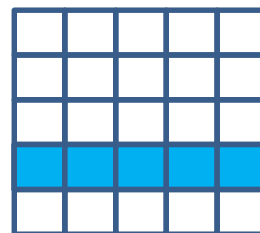
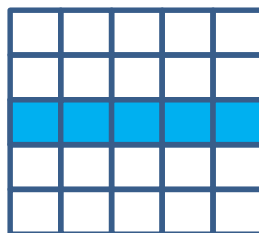
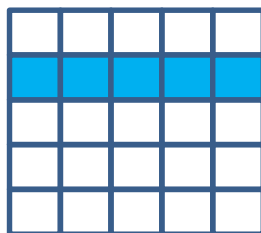
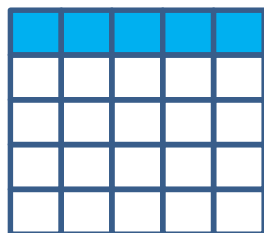


学習データでないと認識されたデータ

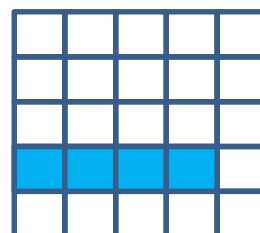
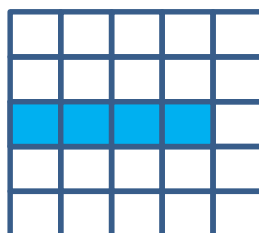
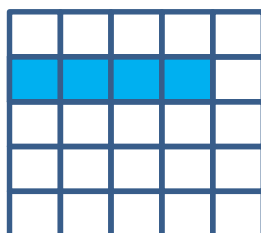
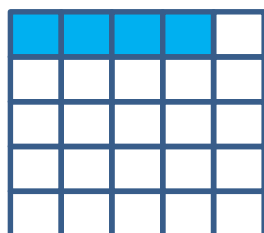


学習用データ(汎化能力のチェック)

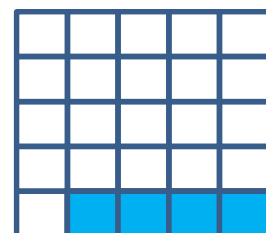
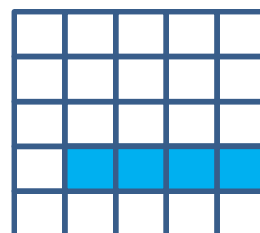
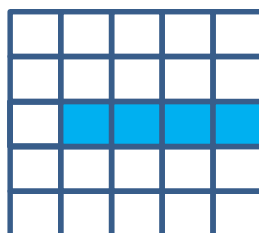
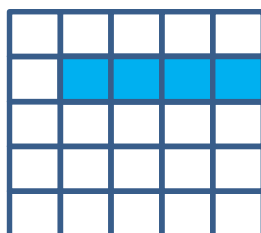
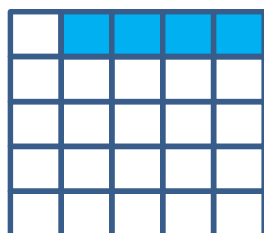
パターン1



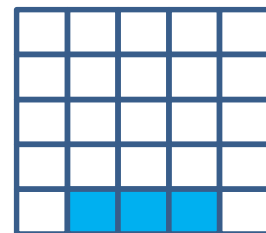
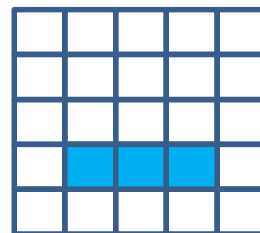
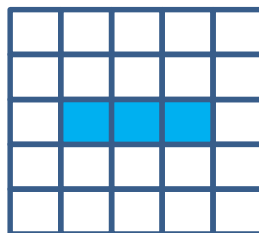
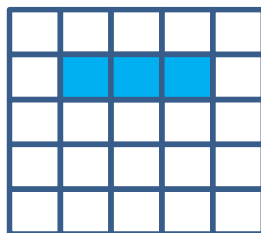
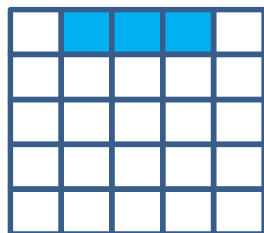
パターン2



パターン3

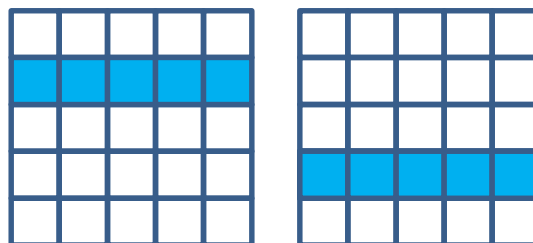


パターン4

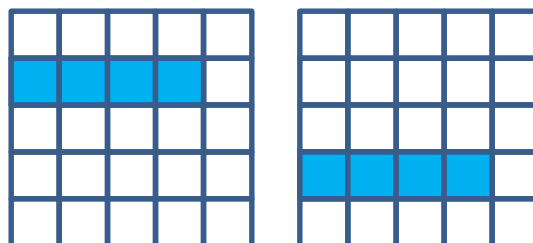


正しく認識されたデータ

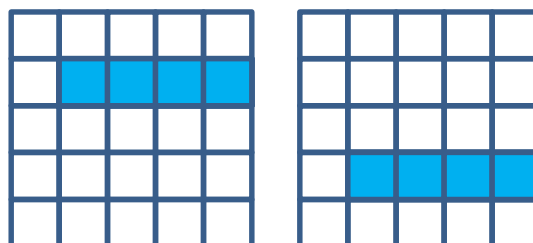
パターン1と認識



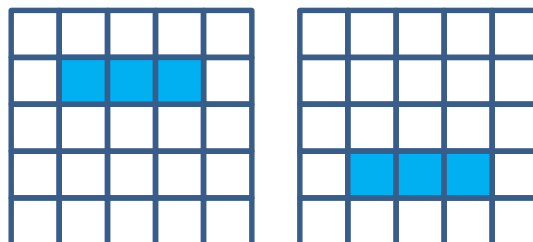
パターン2と認識



パターン3と認識

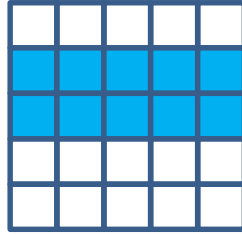


パターン4と認識

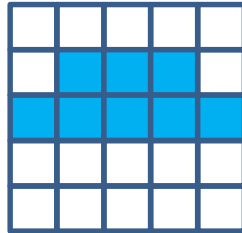


汎化能力チェックのデータ

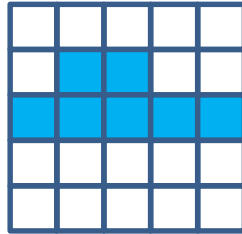
パターン1と認識



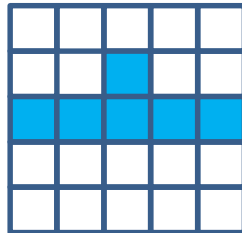
パターン1と認識



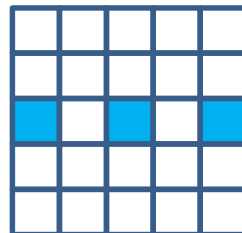
パターン1と認識



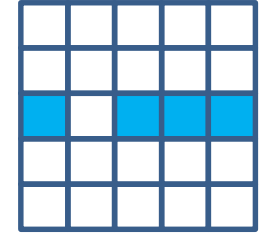
パターン1と認識



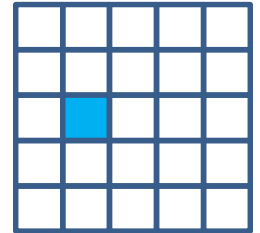
パターン1と認識



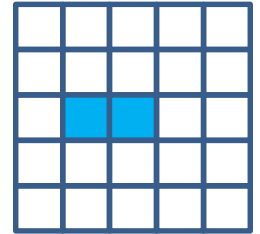
パターン1と認識



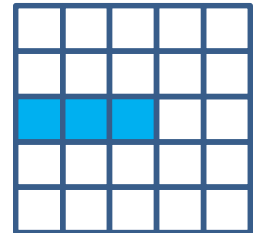
パターン4と認識



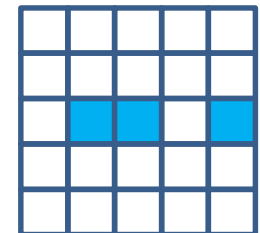
パターン4と認識



パターン2と認識



パターン3と認識



4. 結 言

最近、ディープラーニングによるパターン認識が広く応用されるようになってきた。

しかし、この方法が万能のパターン認識とは言い切れない。

学習に長時間が掛かるとか、過学習による異常な認識結果を生むなどの問題があるが、最大の問題は推論がブラックボックスのため、認識の根拠が分からない点から来る不安であろう。

ベイズ学習は統計学のベイズ推論に基づくものであるから、ディープラーニングとはまったく異なる理論に基づく。

学習方法もまったく異なるので、上述のようなディープラーニングの問題から解放してくれる可能性を有する。

今回の研究は、基本的なアイデアを示したに過ぎない。この方法が、いかなる可能性があるかは今後の研究課題である。