最適なベイズ識別のための状況依存の事前確率の学習

本村 陽一 $^{(PY)}$ (motomura@etl.go.jp)
電子技術総合研究所 情報科学部

Learning Situated Prior Probability for Optimum Bayesian Discrimination

Yoichi Motomura (motomura@etl.go.jp)
Electrotechnical Laboratory Information Science Division

ABSTRACT: For optimum Bayesian discrimination, we have to obtain proper prior probabilities, which sometimes depends on particular situation. These situated prior probabilities can be modeled by Bayesian networks. Especially, Bayesian network on neural networks (BNNN) has some advantages for learning situated prior in the real world environment. As preliminary experiments, the learning situated prior probabilities from an English dictionary is shown. It can apply to handwriting character recognition using Bayesian discrimination. The result improve recognition performance better than an ordinal neural network classifier.

1 はじめに

ベイズ的意志決定理論により人間の意志決定過程を解明する試みがあるが, 実際的な場面における人間の直観的な解との乖離が心理学実験などを通じてしばしば指摘されてきた。しかし, 事前確率 (および損失関数) は一定なものであるのではなく状況や文脈に応じてダイナミックに変化するものと考えることによりこの不整合が解決できるという主張もされている [1]。

パターン認識においてもベイズ事後確率に基づくベイズ識別は理論上最適解を与えるが, 実際には学習時の状況と識別時の状況と差異が著しい場合には識別器が良く学習されていたとしても結果は満足するものとならない。この場合, 識別時の状況固有のパターン発生頻度を状況依存の事前確率として表し, 別途学習した最尤推定に基づく識別器の出力と統合したベイズ解を求めれば, 固定した事前分布を用いるものに比べて性能が向上すると期待できる。

また、ヒトが用いている先験確率を情報処理に利用するために推定しようという試み [2] がある。多様な状況でヒトが自然に使用している先験確率のモデル化のためには文脈や状況 (対象以外の観測データ) に依存したものにすることが必要だろう。すなわち事前確率は対象以外の観測情報から動的に変化する一種の条件付き確率であると考えこれを学習することを考える。この時, 事前確率の表現方法 (モデル) とそれを具体的に実現する学習方法がポイントとなる。

ここでは, 筆者らが提案している神経回路上のベイジアンネット (BNNN: Bayesian network on Neural Networks) を用いて状況に依存した事前確率のモデル化を行なう。これにより明示的な変数レベルで依存関係をグラフ構造で表現し BP 法で学習を行う。

簡単な応用例として英単語辞書により文字の事前確率を学習し, これと各パタンを個別に学習した 4 層の神経回路網と統合して英字認識を行なった結果を示す。

2 状況依存の事前確率を用いるベイズ識別

ラベル C を持つパターン x サンプルセットを学習して識別 (分類) 器を構成する。ここで識別器として多層ニューラルネットを用い, BP 法で学習する方法が良く知られている。理想的な学習後の識別器は近似的に式 (1) の事後確率を最大にするベイズ識別を行なうものとみなせる [3] (各クラスのパタン数を同一にするなどして事前分布を一様分布とした場合は最尤推定と等価となる)。

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{\sum_i P(x|C_i)P(C_i)} \tag{1}$$

実際に識別を行なう場合のパターンの発生頻度が事前分布 $P(C_i)$ と合致している場合にはこのベイズ識別は理論上最適であるが, 実際には識別時の状況が学習時の状況と一致しない可能性が生じる。例えばあるパタン x をクラス C_i のどれかに分類する時, 実際にはクラス間の発生頻度が観測時間や観測場所に依存しているような場合, 事前分布 ($P(C_i)$) は状況 (S) に依存したものになっている。そこでこれを条件付き確率 $P(C_i|S)$ として考え, 別の確率モデルを用いて学習する。これと通常の最尤解を学習した識別器から得られる尤度 $P(x|C_i)$ とを統合した式 (2) を最大とするクラスを識別結果とする。

$$P(C_i|x, S) = \frac{P(x|C_i)P(C_i|S)}{\sum_i P(x|C_i)P(C_i|S)} \tag{2}$$

3 ベイジアンネット

前記の条件付き確率 $P(C_i|S)$ を表すためにベイジアンネットを用いると, 複雑な依存関係にある確率変数の間の関係を有向グラフで与えることで様々な状況に応じた条件付き確率をモデル化することができる。特に条件付き確率を神経回路網で学習する BNNN を用いると連続値, 多次元ベクトルを含むような多様な確率変数の取扱いが容易になるメリットがある [8, 9]。このモデルは 3 層の神経回路を多段に接続した階層型神経回路網の上に仮想的にベイジアンネットが表現されたものになっている。例えば確率変数 A, B, C, D の同時確率が

$$P(A, B, C, D) = P(D|C)P(C|A, B)P(B)P(A)$$

のような関係で表せる場合, 図 1の左側のようなグラフでモデル化し, 各変数の条件付き確率を学習した同図右側のニューラルネットにより実現できる。状況に依存する事前確率は状況 S をいくつかの確率変数の組で記述した条件付き確率をベイジアンネットで記述できれば, 十分な実データを用いて学習できる。

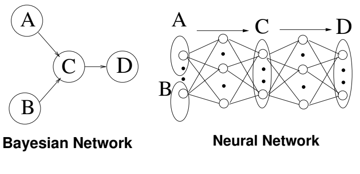


Fig. 1: 神経回路上のベイジアンネット (BNNN)

Key Words: 情報統合, ベイズ識別, 確率推論

4　シミュレーション

ここでは状況に依存する事前確率の学習と有効性を示すための比較的簡単な例題として手書き文字認識を考える。認識モジュールとして 20x20 のドットマトリクス上にマウスで書いた手書き英字 10 文字 {A,B,C,D,E,H,I,J,K,T} を 2x2 の受容野を持つ入力素子 100 個, 中間素子 10 個の隠れ層 2 層, 出力素子 10 個の 4 層の神経回路網, $g(x)$ を用いて平均二乗誤差が 0.01 以下になるまで学習した。この出力を文字パタン x を与えた時の文字 i に対応する近似尤度 $f_i(x)$ と書く。

一方, 状況に依存する事前確率として, 隣接する英字 C^{-1}, C^{+1} が与えられた時の間に挟まれた英字 C^0 に関する条件付き確率 (26 次元確率ベクトル) $P(C^0|C^{-1}, C^{+1})$ を考え, 4 万 3 千語からなる英単語辞書のデータを教師信号として BNNN(この場合, 54-10-26 の神経回路網) を学習した。$S = (C^{-1}, C^{+1})$ を与えた時の C_0 の 26 次元確率ベクトルのうち文字 i に対応するものを $g_i(S)$ と書く。

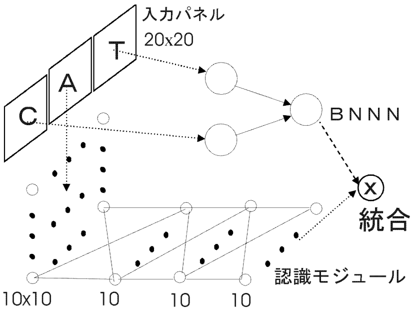
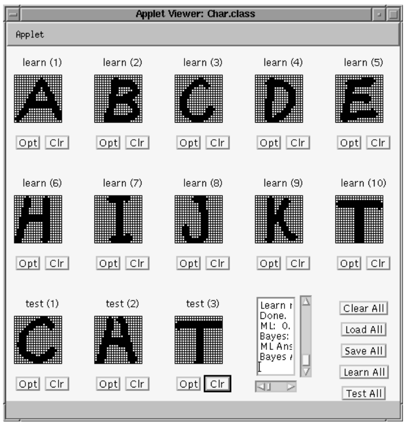


Fig. 2: 実験システム概要



Learn r
Done.
ML: 0.
Bayes:
ML Ans
Bayes /

Fig. 3: 上二段:学習パタン, 下一段:テストパタン

学習後に判読の困難な認識対象 (図 4の最下段) を与え, 次の I.(最尤解),II.(式 (2) に基づくベイズ解) の二者の比較を行なう。

I.$P(x|C_i) = f_i(x)/\sum_i f_i(x)$

II.$P(C_i|x, S) = P(x|C_i) \cdot g_i(S)/\sum_i(P(x|C_i) \cdot g_i(S))$

C^{-1} ='C', C^{+1} ='T' と認識した時の 2 文字目に関する I., II. の推定結果を下表 (単位:%) に示す。

	A	B	C	D	E	H	I	J	K	T
I	15.6	3.3	17.9	9.1	7.5	22.1	8.6	3.8	9.8	2.1
II	75.2	0.0	0.4	0.0	3.3	4.7	12.7	0.0	3.8	0.0

これにより I. では誤認識されていた 2 文字目の判読困難な文字が, II. では ‘A’ と認識できている。

5　議論

前後の依存関係を考慮して学習する認識モデルとして HMM(Hidden Markov) があり, 手書き認識へも適用されている [4] が, HMM では識別する単語毎にモデル化するのに対し, 本手法では文脈依存性と個別の最尤識別学習を独立に学習し, 識別時に統合する点が大きく異なっている。一方, 選択的注意機構のモデル [5] では認識モジュールと高次情報の制御回路が相互により密接に関わっており, このような認識過程のダイナミクスにおける事前知識の利用についても引続き検討を行ないたい。また記号とパターンの統合 [6, 7] という観点では, BNNN は任意の確率変数の相互関係についてはより抽象度の高いグラフ構造で表現する特長がある。今後は我々が身近に接しているより広範な問題について, 文脈依存関係を自然にモデル化し認知的にもより人間に近い振舞を実現することが課題である。

6　謝辞

ホームページのアプレットを参考にさせていただいた慶応大学成田氏に感謝いたします。

参考文献

[1] 繁桝: “意志決定の認知統計学”, 朝倉書店, (1995).

[2] 上坂, 山岸, 野沢: “学習におけるヒトの予備知識について”, 神経回路学会第 6 回全国大会論文集, p.243–244, (1995).

[3] 大津: “パターン認識と学習の基礎理論”, ニューロコンピューティングとその周辺, 脳とコンピュータ, 2, 培風館, (1995).

[4] Se-Chang Oh, Jin-Young Ha and Jin-H.Kim: “Context Dependent Search in Interconnected Hidden Markov Model for Unconstrained Hand Writing Recognition”, *Pattern Recognition*, Vol.28, No.11, pp.1693–1704, (1995).

[5] 庄野, 福島: “選択的注意機構を用いた英文筆記体文字列認識”, 神経回路学会第 6 回全国大会論文集, p.272–273, (1995).

[6] 石川, 松尾: “視覚認知におけるコンテクスト依存性の統合的モデル化の試み”, 記号とパターンの統合処理ワークショップ論文集, pp.9–14, (1994).

[7] 西, 大関, 桜井, 大森: “記号とパターンの統合による画像理解のための一方式”, 計測自動制御学会論文誌, vol.30, No.11, pp.1368–1377, (1994).

[8] Y,Motomura, I.Hara, H.Asoh, and T.Matsui: “Bayesian Network that Learns Conditional Probabilities by Neural Networks” , *Proc. of International Conference on Neural Information Processing '97* ,(1997).

[9] 本村 “多層パーセプトロンを用いたベイジアンネットによる確率推論”, 神経回路学会第 7 回全国大会論文集, p.243–244, (1996).