

脳内ダイポール個数推定のためのLVQの学習について

5N-8

北島伸克 上條憲一(NEC 情報メディア研究所)

1 はじめに

LVQ(Learning Vector Quantization, 学習ベクトル量子化)[3]は、幅広い応用分野に適用可能な、高速・高性能のカテゴリ分類法であるといわれている。しかしながら、正答率に影響する様々なパラメタの設定法が明確でないという問題点がある。本稿では、LVQの応用例として脳内ダイポール個数推定を取り上げた。まずはじめに、トレーニングデータの分布に適合した参照ベクトルの初期値設定法を新たに提案したので、その有効性を示す。さらに、実験により参照ベクトル数や学習回数と正答率との関係を調べたので、その結果を述べ、考察を加える。BP法との比較についても触れる。

2 LVQを適用するときの問題点

LVQを用いる際に考慮した項目は以下の通りである。

1. 参照ベクトルの初期値。

LVQの正答率は、参照ベクトルの初期値によってかなり影響を受けるといわれている[4]。参照ベクトルの初期値がトレーニングデータの分布から外れないこと、トレーニングデータ分布に適合していること等が高い正答率につながると思われる。

2. 参照ベクトル数。

高い正答率を得るための適切な参照ベクトル数を決定するために、参照ベクトル数と正答率の関係を調べる必要がある。参照ベクトル数の増加に伴って、ある程度までは正答率が向上すると予想できるが、増加すればするほど正答率が向上するという保証はない。参照ベクトル数は学習時間に直接影響するため、同程度の正答率の場合には参照ベクトル数は少ない方が好ましい。

3. 学習回数。

適切な学習回数を決めるために学習回数と正答率の関係を調べる必要がある。学習回数も、参照ベクトル数同様に、学習時間を短くするためには、少ない方が好ましい。

3 脳内ダイポール個数推定への応用

LVQ1[3]を脳内ダイポール個数推定に適用した。脳内ダイポールによる脳内活動源のモデル化は、脳機能の定量的な解析の一方法である[1]。この方法では、頭皮上で計測された電磁場から脳内の仮想的なダイポールの位置と個数の推定を行う。個数推定に注目すると、そのデータ[2]は、電極数個分の電位差の組とダイポールの個数を示すラベルから構成されている。よって、電極数(30)個の電位差の組をベクトルと解釈し、ラベルをカテゴリと見なすと、ダイポールの個数推定は電極数個の次元のデータ空間をダイポールの個数に対応するカテゴリに分類する問題ととらえることができる。よって、LVQを用いた脳内ダイポール個数推定が可能である。全ての実験に共通の設定を以下に示す。

- トレーニングデータ数: 5,000個
- 学習係数の初期値 $\alpha(0)$: 0.01
- 学習係数の更新式: $\alpha(n) = \alpha(0)(1 - n/N)$, ただし N は全学習回数, n は現在までの学習ステップ数。

3.1 参照ベクトルの初期値設定

今回、次の3種類の方法を用いて参照ベクトルの初期値設定を行った。

1. トレーニングデータと絶対値がほぼ同じ大きさの乱数を初期値にした。
2. トレーニングデータの一部をランダムに取り出したものを初期値にした。
3. 2を自己組織化[3]したものを初期値にした。

1. はトレーニングデータを大きくはずれないように考慮した1つの方法であり、2. は1.をより確実にして、参照ベクトルがトレーニングデータの分布内に必ず存在するようにした方法である。3. は参照ベクトルをトレーニングデータ空間全体の分布に適合させることが正答率の向上に役立つであろうという予測に基づいて考案した方法である。3種類の初期値設定法による結果を表1に示す。結果を見ると予想通り1, 2, 3と順を追うごとに正答率が向上している。2が1より良い結果が出ている

表 1: 3種類の初期値設定法による正答率

初期値設定法	正答率
1. ランダム法	74.5%
2. トレーニングデータ法	79.8%
3. 自己組織化法	80.5%

表 2: 参照ベクトル数による正答率の変化

参照ベクトル数	70	100	200	400	1,000
正答率 (%)	78.5	80.5	80.7	85.5	83.5

のは、推定に貢献しない参照ベクトルが減少したからであると思われる。また、3でさらによい結果が出ているのは、参照ベクトルの配置がトレーニングデータの分布に従っていることが正答率の向上に貢献しているからであると思われる。

3.2 参照ベクトル数と学習回数

参照ベクトル数と正答率の関係を調べる実験を行った。結果を表2に示す。参照ベクトル数と正答率は比例関係にはなく、正答率に対して適切な参照ベクトル数が存在することがわかった。

このときの学習回数は、適用したトレーニングデータとそれに最近接の参照ベクトルの間の距離の2乗の和の学習回数に対する変化(以後、収束曲線と呼ぶ)から判断した(図1)。その結果学習回数と正答率も比例関係にはなく、正答率に対する適切な学習回数が存在することが判明した。

3.3 BP法との比較

ニューラルネットの1手法であるBP法を用いてダイボールの個数推定が行われている[2]。LVQの性能を検証するため、BP法との比較を表3に示す。

LVQによる結果とBP法による結果を比較すると、

表 3: LVQとBPの学習時間、推定時間、正答率の比較

	LVQ	BP
学習時間	約3時間9分	約24時間
推定時間	約2.5秒	約0.007秒
正答率 (%)	85.5	89.4

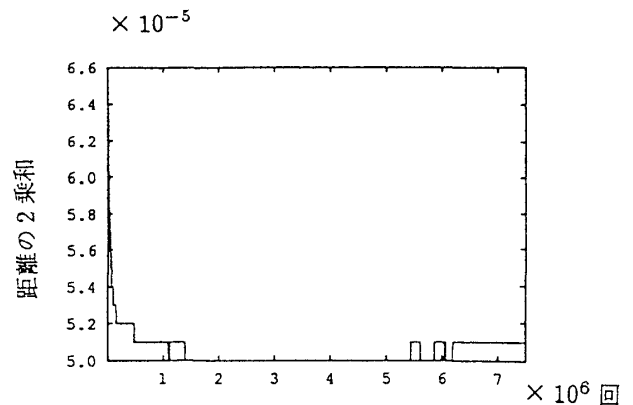


図 1: 収束曲線の例, 横軸: 学習回数 ($\times 10^6$ 回), 縦軸: 適用したトレーニングデータとそれに最近接の参照ベクトルの間の距離の2乗和

現時点ではLVQを用いた方法は正答率ではやや劣るものの、学習はかなり高速化されている。また、正答率に関しても自己組織化の適用に工夫を加えるなどすれば向上の余地があると思われる。

4 まとめ

LVQを脳内ダイボール個数推定に適用した。正答率に対する参照ベクトルの初期値の影響が大きいことがわかった。自己組織化を用いて、トレーニングデータの分布に適合した参照ベクトルの初期値を設定する方法を提案した。参照ベクトル数や学習回数は正答率とは比例関係にはなく、正答率に対する最適な参照ベクトル数や学習回数が存在することを示した。LVQを用いることによりBPに比べてかなり高速に学習できることがわかった。

参考文献

- [1] Bin He et al.: *IEEE Trans. Biomed. Eng.*, Vol. BME-34, No.6, pp.406-414, 1987.
- [2] 喜友名朝春, 谷川哲司, 上條憲一: 第7回生体・生理工学シンポジウム論文集, Vol.7, pp.415-418, 1992.
- [3] Kohonen, T.: *Proc. IEEE*, Vol.78, No.9, pp.1464-1480, September, 1990.
- [4] 横田まなみ, 片桐滋, エリック・マクダーモット: 日本音響学会講演論文集, 2-P-29, pp.287, 1989.3.