

CCCPを用いた最大マージンクラスタリングのカーネル化

塚原 裕史 †

高橋 宏子 ‡

趙 晋輝 ††

†(株)デンソーアイティラボラトリ

‡中央大学大学院理工学研究科

††中央大学理工学部情報工学科

1 はじめに

最大マージンクラスタリング (MMC) [1] は、優れた性能を持つことが示されている [2] が、一方で目的関数の非凸性により、計算コスト的な課題がある。これまでに、MMC を凸問題へ緩和して高速に近似解を求める方法が研究されてきた。例えば、半正定値計画問題に緩和する方法 [3]、サポートベクトルマシンを逐次的に解く手法 [4] などがある。しかし、前者ではなおも計算コストが高く、大規模データへの適用は困難であり、後者では収束性に関する保証がないなどの問題がある。一方、Zhao らは、convex-concave procedure (CCCP) と切断平面 (CP) アルゴリズムを用いた方法を提案し、他の手法に比べて、高速かつ優れたクラスタリング性能が得られることを示した [5]。但し、このアルゴリズムにカーネル法を適用する場合、カーネル行列のコレスキー分解を行う必要があり、現実的な問題に適用する場合には、数値的な困難がある。本研究では、彼らの方法をカーネル行列のコレスキー分解を行うことなく実行できるように変形したアルゴリズム [6] によって、照明変化によって作られる顔画像の部分空間をグラスマンカーネル [7] を用いてクラスタリングを行い、提案アルゴリズム [6] の有効性を検証した結果について述べる。

2 アルゴリズム

Xu らによる MMC の定式化 [1] では、目的関数がデータのラベルという離散的な変数に関して非凸性を持ち、それらに関して組合せ最適化を行う必要がある点に課題があった。Zhao らは、ラベル変数を含まず、かつ MMC と等価な最適解を与える最適化問題を提案した [5]。この問題では、目的関数は凸関数であるが、制約条件が特異性を有する非線形関数で与えられる。そのため、CCCP を適用し、凸二次計画問題を逐次解くことで最適解の近似解を高速に求める方法が提案された。但し、この方法にカーネル法を適用する場合、カーネル行列のコレスキー分解を求めなければならないが、我々は次のような

カーネル関数を陽に用いることができるように変形したアルゴリズムを提案した [6]: (本論文では、CP アルゴリズムは適用していないことに注意する。)

$$\max_{\lambda_i, \mu_1, \mu_2} L(\{\lambda_i\}, \mu_1, \mu_2), \quad (1)$$

$$\text{s.t. } 0 \leq \lambda_i \leq \frac{C}{n}, \quad (i = 1, 2, \dots, n) \quad (2)$$

$$n(\mu_1 - \mu_2) - \sum_i s_{i,t} \lambda_i. \quad (3)$$

但し、ここで、補助変数 $\{s_{i,t}\}$ を導入した。また、 n はデータ数、 l はクラスタのサイズ比を調整するパラメタ、 C はソフトマージンによるパラメタである。さらに、目的関数 $L(\{\lambda_i\}, \mu_1, \mu_2)$ は、以下で与えられる:

$$\begin{aligned} L(\{\lambda_i\}, \mu_1, \mu_2) = & -\frac{1}{2} \sum_{i,j} s_{i,t-1} s_{j,t-1} K(\mathbf{x}_i, \mathbf{x}_j) \lambda_i \lambda_j \\ & + \sum_{i,j} s_{i,t-1} K(\mathbf{x}_i, \mathbf{x}_j) \lambda_i (\mu_1 - \mu_2) - \frac{1}{2} \sum_{i,j} K(\mathbf{x}_i, \mathbf{x}_j) (\mu_1 - \mu_2)^2 \\ & - (\mu_1 + \mu_2) l + \sum_i \lambda_i. \end{aligned} \quad (4)$$

この提案方法では、変数 $\{\lambda_i\}, \mu_1, \mu_2$ の最適化と補助変数の更新とを交互に実行する。(1)–(3)の解を $\{\lambda_{i,t+1}\}, \mu_{1,t+1}, \mu_{2,t+1}$ とすると、判別関数は

$$f(\mathbf{x}) = \sum_i (s_{i,t} \lambda_{i,t+1} - \mu_{1,t+1} + \mu_{2,t+1}) K(\mathbf{x}_i, \mathbf{x}) + b_{t+1}, \quad (5)$$

で与えられる。但し、ここでバイアス項 b_{t+1} は、KKT 条件から $0 < \lambda_{i,t+1} < C/n$ を満たす i に対して、

$$b_{t+1} = s_{i,t} + \sum_j K(\mathbf{x}_i, \mathbf{x}_j) (\mu_{1,t+1} - \mu_{2,t+1} - s_{i,t} \lambda_{j,t+1}), \quad (6)$$

で求められる。補助変数 $s_{i,t}$ は、データ x_i の判別関数値の符号に相当し、次の式に従って更新する:

$$\begin{aligned} s_{i,t+1} = & \text{sgn} \left(\sum_j \lambda_{j,t+1} s_{j,t} K(\mathbf{x}_i, \mathbf{x}_j) \right. \\ & \left. - (\mu_{1,t+1} - \mu_{2,t+1}) \sum_j K(\mathbf{x}_i, \mathbf{x}_j) + b_{t+1} \right). \end{aligned} \quad (7)$$

以下に示す手順を繰り返し、その収束解と判別関数 (5) を用いて、クラスタリングを実行する。

Kernelization of convex-concave procedure relaxation of maximum margin clustering

†Hiroshi Tsukahara ‡Hiroko Takahashi ††Jin Hui Chao

†DENSO IT LABORATORY, INC.

‡Graduate School of Science and Engineering, Chuo University

††Faculty of Science and Engineering, Chuo University

アルゴリズム

- 1: $s_{i,0} = 0$ とする.
 - 2: $s_{i,t}$ を用いて, 問題 (1)–(3) を解き, $\{\lambda_{i,t+1}\}, \mu_{1,t+1}, \mu_{2,t+1}$ を求める.
 - 3: (6) により b_{t+1} を求める.
 - 4: (7) により, $s_{i,t+1}$ を更新する.
 - 5: 目的関数 (4) の値が収束するまで, 2–4 のステップを繰り返す.
-

3 評価実験

図.1,2 に, Extended Yale Face B Datasets [8] のデータを利用し, 照明変化によって作られる顔画像 (正面向き) の部分空間 (以下, 顔照明空間) をクラスタリングした結果を示す. 各顔照明空間は 64 枚 (照明条件) の画像から成る. ここから 40 枚のセットを 100 組ランダムに抽出し, 各顔画像の画素値を縦横に射影したヒストグラムをベクトル化し, それらを照明条件の違いで並べた行列に非負行列因子分解を適用し, 顔照明空間にフィットする k 次元部分空間の基底を求め, 求められた基底に対して, グラスマンカーネル [7] を適用した. 各図の左右に, カーネル行列の値と判別関数の値を示した. 図.1 は, 2 次元部分空間でフィットした結果の例である. 2 次元のような低次元でも比較良いクラスタリング結果が得られており, 平均誤識別率は約 4.5% 程度であった. 今回, 図.2 に示すように, 5 次元の場合にもっとも良い結果が得られ, 平均誤識別率は 0.05% 程度と, ほぼ正確なクラスタリング結果が得られた.

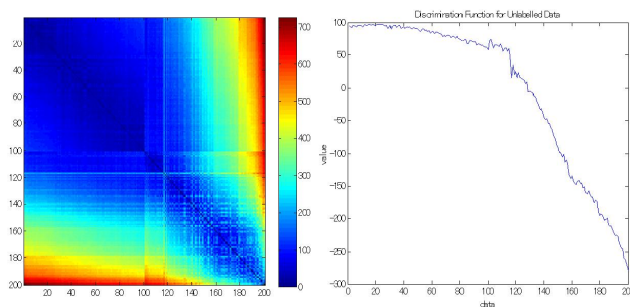


図 1: 2 次元部分空間でのクラスタリング結果

4 まとめ

カーネル関数を陽に用いて高速に MMC を実行するアルゴリズムを提案し, グラスマンカーネルという非自明なカーネルを用いた顔照明画像のクラスタリングという実際的な問題に提案手法を適用し, その有効性を確認した. 今後, 半教師付き問題への拡張を検討する.

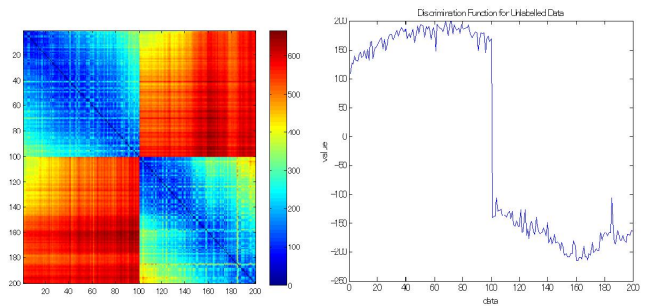


図 2: 5 次元部分空間でのクラスタリング結果

参考文献

- [1] L. Xu, J. Neufeld, B. Larson, and D. Schuurmans. Maximum margin clustering. In *Advances in Neural Information Processing Systems*, Vol. 17, pp. 1537–1544, 2004.
- [2] O. Chapelle, V. Sindhwani, and S. S. Keerthi. Branch and bound for semi-supervised support vector machines. In *Advances in Neural Information Processing Systems*, Vol. 19, pp. 217–224, 2006.
- [3] H. Valizadegan and R. Jin. Maximum margin clustering and unsupervised kernel learning. In *Advances in Neural Information Processing Systems*, Vol. 19, pp. 1417–1424, 2006.
- [4] K. Zhang, I. W. Tsang, and J. T. Kowk. Maximum margin clustering made practical. In *Proceedings of the 24th International Conference on Machine Learning*, 2007.
- [5] B. Zhao, F. Wang, and C. Zhang. Efficient maximum margin clustering via cutting plane algorithm. In *ICML*, pp. 751–762, 2008.
- [6] 高橋宏子, 塚原裕史, 趙晋輝. 最大マージンクラスタリングにおける cutting-plane アルゴリズムへのカーネル法の適用. 日本神経回路学会第 19 回全国大会, 2009.
- [7] P. Gruber and F.J. Theis. Grassmann clustering. In *EUSIPCO*, 2006.
- [8] K.C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence*, Vol. 27, No. 5, pp. 684–698, 2005.