1F5-1in1

潜在変数を用いたカーネルの確率変数化による類似度からのクラスタリング

Clustering with Affinities by Kernel Function as Random Variable

竹岡 邦紘 *1　　岡留 剛 *1
Kunihiro Takeoka　　Takeshi Okadome

*1関西学院大学大学院理工学研究科
Graduate School of Science and Technology, Kwansei Gakuin University

The method proposed here identifies clusters of entities with their affinities by the posterior maximization on the basis of a generative model represented by a kernel function as a random variable, where for every entity, there is a corresponding M-dimensional binary latent variable governed by a Bernoulli distribution with parameter specific to the cluster to which an entity belongs. Because of intractability of the optimization, an approximation is obtained by the following: (1) for a selected row of the similarity matrix, the entities are divided into two groups by the posterior maximization, (2) the same procedure continues until the entities are not divided into two groups, and (3) the entities are reclustered by the K-nearest-neighbour technique using clustering results as labels. The evaluation using the MNIST dataset reveals that the performance using the proposed method is equal to that by the methods developed in the previous studies.

1. はじめに

データ間の類似度が与えられる離散構造データに対しては，一般的なクラスタリング手法である K-means 法などを利用することができない．

類似度データに対するクラスタリング手法の代表例として，スペクトラルクラスタリングがある．[Shi 00] で初めて提案され，[Ng 02] などによって一般化された手法である．スペクトラルクラスタリングは，個体間の類似度行列についてある固有値問題を解くことで，各個体を特徴ベクトルで表現した行列に変換する手法である．その変換された行列においては，通常の数値データに対するクラスタリング手法が利用できるため，K-means などの一般的な手法を用いてクラスタリングを行う．その手法においては，[Rasmussen 00] を用いることでクラス数が未知でもクラスタリングを行うことができる．その手法については，[Socher 11] に詳しい．

本稿では，類似度行列が確率的生成モデルに従って生成されると考え，事後確率最大化によってクラスタリングを行う手法を提案する．

2. 類似度データに対する確率的生成モデル

2.1 問題の定式化

いくつかのクラスに分かれる N 個の個体間の類似度行列 $\mathbf{K}$ が与えられたとする．すなわち，個体には 1 から N まで番号がふられ，個体 i と個体 j の類似度が行列 $\mathbf{K}$ の k_{ij} 要素である．$\mathbf{K}$ の各要素 $k_{ij} = k_{ji},\ i, j = 1, ..., N,$ は整数とする．

2.2 類似度データの生成モデル

各個体は，それぞれ M 次元の潜在変数 $\boldsymbol{z}_i = (z_{i1}, z_{i2}, ..., z_{iM})^{\mathrm{T}},\ z_{im} \in \{0, 1\},\ m = 1, ..., M,\ i = 1, ..., N,$ を持つ．ただし，M は正の整数をとるパラメータであり以下を満たすとする．

$$0 \quad \leq \quad k_{ij} \leq k_{ii}\ (k_{jj}) \leq M. \tag{1}$$

また，潜在変数の各要素は，クラスごとに定まった平均を持つ独立なベルヌーイ分布に従うとする．すなわち，

$$z_{im} \sim Bern(q_c),\ m = 1, .., M\ (c \text{ は個体 } i \text{ が所属するクラス}).$$

この潜在変数を用いて，$\mathbf{K}$ の要素は通常の内積として表現されると仮定する．すなわち，

$$k_{ij} = \boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{z}_j. \tag{2}$$

式 (1) は，すべての個体は自身との類似度が最大であり，その最大値が M であることを意味する．式 (2) により，類似度行列の要素が特徴空間での線形カーネルによって定義され，これは，カーネル関数の潜在変数を用いた確率変数化に相当する *1．以下では，クラス数を C とし，$\boldsymbol{q} = (q_1, q_2, ..., q_C)^{\mathrm{T}}$ を，潜在変数を生成するベルヌーイ分布のパラメータ（クラスごとの平均）とする．また，1-of-C 符号化法で個体 i が属するクラスを表す潜在変数 $\boldsymbol{s}_i = (s_{i1}, s_{i2}, ..., s_{iC})^{\mathrm{T}},\ s_{ic} \in \{0, 1\},\ \sum_{c=1}^{C} s_{ic} = 1,$ を導入し，$\mathbf{S} = (\boldsymbol{s}_1, ..., \boldsymbol{s}_N)^{\mathrm{T}}$ とする．$\mathbf{S}$ は $N \times C$ 行列である．

類似度行列 $\mathbf{K}$ を対角要素 k_{ii} と非対角要素 k_{ij} に分けて考える．類似度の生成モデルから任意の対角要素 k_{ii} の生成確率は以下となる．

$$p(k_{ii}|\boldsymbol{s}_i, \boldsymbol{q}, M) = {}_M C_{k_{ii}} \prod_{c=1}^{C} (q_c^{k_{ii}} (1 - q_c)^{M - k_{ii}})^{s_{ic}}. \tag{3}$$

ここで，${}_M C_{k_{ii}}$ は二項係数である．類似度の生成モデルより，類似度行列 $\mathbf{K}$ のすべての対角要素は独立となり，すべての対角要素の尤度はそれらの積として定まる．また，非対角要素はすべての対角要素で条件付けたもとで条件付き独立であることが証明できる．非対角要素 k_{ij} が取りうる値は対角要素 k_{ii} と k_{jj} に依存し，潜在変数 $\boldsymbol{z}_i$ と $\boldsymbol{z}_j$ の実際の値をとるビットのパターン数と，取りうる可能性すべてのパターン数の比として

連絡先: 竹岡邦紘，関西学院大学大学院理工学研究科，kunihiro.takeoka@kwansei.ac.jp

*1 集合 $\mathcal{X} \times \mathcal{X}$ 上のカーネル関数 $k(x, y)$ に対応する再生核ヒルベルト空間を $\mathcal{H}_k$ としたとき，$\mathcal{X}$ 上の確率変数 x に対して特徴写像 $k(\cdot, X)$ は $\mathcal{H}_k$ に値をとる確率変数となる [Berlinet 04]．それに対し，本研究では，X, Y を $\mathcal{X}$ 上の確率変数としたときの $k(X, Y)$ を考え，これは $\mathbb{R}$ に値をとる確率変数となる．

生成確率を計算することができる．これより，非対角要素 k_{ij} の尤度は以下となる．

$$p(k_{ij}|k_{11},...,k_{NN},M) = \frac{{}_{k_{ii}}C_{k_{ij}} \cdot {}_{M-k_{ii}}C_{k_{jj}-k_{ij}}}{\sum_{x_{ij}=\max(0,k_{ii}+k_{jj}-M)}^{\min(k_{ii},k_{jj})} {}_{k_{ii}}C_{x_{ij}} \cdot {}_{M-k_{ii}}C_{k_{jj}-x_{ij}}} \quad (4)$$

したがって，式 (3) と (4) から類似度行列 $\mathbf{K}$ の尤度は以下となる．

$$\begin{aligned} p(\mathbf{K}|\mathbf{S},\boldsymbol{q},M) &= \prod_{i=1}^{N} p(k_{ii}|\boldsymbol{s}_i,\boldsymbol{q},M) \prod_{j=i+1}^{N} p(k_{ij}|k_{11},...,k_{NN},M) \\ &= \prod_{i=1}^{N} {}_{M}C_{k_{ii}} \prod_{c=1}^{C} (q_c^{k_{ii}}(1-q_c)^{M-k_{ii}})^{s_{ic}} \\ &\times \prod_{i=1}^{N}\prod_{j=i+1}^{N} p(k_{ij}|k_{11},...,k_{NN},M), \end{aligned}$$

$$p(k_{ij}|k_{11},...,k_{NN},M) = \frac{{}_{k_{ii}}C_{k_{ij}} \cdot {}_{M-k_{ii}}C_{k_{jj}-k_{ij}}}{\sum_{x_{ij}=\max(0,k_{ii}+k_{jj}-M)}^{\min(k_{ii},k_{jj})} {}_{k_{ii}}C_{x_{ij}} \cdot {}_{M-k_{ii}}C_{k_{jj}-x_{ij}}}.$$

一般に，類似度行列の要素間に相関があり,「実効的」なデータ数は類似度データ数の $\frac{N(N+1)}{2}$ よりもかなり少ない．そのため，尤度最大でのパラメータ学習では過学習が起こると考えられる．しかし，この生成モデルのもとで，事前確率を設定し，事後確率が最大となるパラメータを直接求めることが望ましいが，実際上計算が困難である．特に，M が非対角要素の条件付き確率に複雑に影響し，M の事後確率を計算することは困難である．そこで，以下の方略が考えられる．(1)M をモデルパラメータとし，$\mathbf{S}$ と $\boldsymbol{q}$ に関して積分消去してベイズモデル比較を行う．(2)M を確率変数として扱い，その事前分布を導入し，事後確率の最大値をサンプリングにより求める．(3) 行列 $\mathbf{K}$ を行ごとに分割して考える．この場合は，後で示すように M の値はクラスタリングと無関係になる．本稿では，次の章において (3) について詳述する．

3. 行分割による近似

類似度行列を行ごとに分割し，事後確率最大化の計算を行う近似手法を述べる．すなわち，(1) 類似度行列を行ごとに分割し，各行ごとに類似度行列 $\mathbf{K}$ の対角要素で条件付けたもとでの事後確率の最大化を行い，対角要素を含むクラス A とそれ以外のクラス B とする 2 クラスに分割し，(2) クラス B の各要素について，他の行を用いて同様のことを行い，(3) これをすべての個体の所属クラスが決まるまで繰り返す．最後に，(4)K 近傍法を用いて個体の所属クラスの更新する．

3.1 一つの行に着目した 2 分割

類似度行列 $\mathbf{K}$ の一つの行に着目する．類似度行列 $\mathbf{K}$ の第 m 行 $\boldsymbol{k}_m$ は，個体 m と，すべての個体との類似度を要素とするベクトルである．この第 m 行の対角要素で条件付けられた尤度は以下となる．

$$\begin{aligned} p(\boldsymbol{k}_{-m}|k_{mm},\mathbf{S},\boldsymbol{q}) &= \prod_{i=1}^{N} p(k_{mi}|k_{mm},\boldsymbol{s}_i,\boldsymbol{q}) \\ &= \prod_{i=1}^{N} {}_{k_{mm}}C_{k_{mi}} \prod_{c=1}^{C} (q_c^{k_{mi}}(1-q_c)^{k_{mm}-k_{mi}})^{s_{ic}}. \end{aligned}$$

$\boldsymbol{k}_{-m}$ は類似度行列 $\mathbf{K}$ の第 m 行である $\boldsymbol{k}_m$ から k_{mm} のみを除いた $N-1$ 次元ベクトルである．行ごとの尤度の式から以下がわかる．すなわち，(1) 各行について，対角要素で条件付けた非対角要素は独立となる．(2) 第 m 行に関する尤度には潜在変数の次元数 M が現れない．

$$\begin{aligned} p(\boldsymbol{q}|\boldsymbol{\alpha},\boldsymbol{\beta}) &= \prod_{c=1}^{C} \frac{\Gamma(\alpha_c+\beta_c)}{\Gamma(\alpha_c)\Gamma(\beta_c)} q_c^{\alpha_c}(1-q_c)^{\beta_c}. \\ p(\mathbf{S}|\boldsymbol{\pi}) &= \prod_{i=1}^{N}\prod_{c=1}^{C} \pi^{s_{ic}}. \\ p(\boldsymbol{\pi}|\boldsymbol{\gamma}) &\propto \prod_{c=1}^{C} \pi_c^{\gamma_c-1}. \end{aligned}$$

潜在変数とパラメータの事後確率を求めるために，共役な事前分布を導入する．すなわち，$\boldsymbol{q}$ の事前分布はベータ分布とし，その超パラメータをそれぞれ $\boldsymbol{\alpha}$,$\boldsymbol{\beta}$ とする．$\mathbf{S}$ の事前分布はカテゴリカル分布とし，そのパラメータを $\boldsymbol{\pi}$ とする．$\boldsymbol{\pi}$は C 次元のベクトルであり,$\sum_{c=1}^{C}\pi_c=1$ という制約を持つ．さらに，$\boldsymbol{\pi}$ の事前分布はディリクレ分布とし，その超パラメータを $\boldsymbol{\gamma}$ とする．

これらより，第 m 行についての潜在変数とパラメータに関する事後確率は以下となる．

$$\begin{aligned} p(\mathbf{S},\boldsymbol{\pi},\boldsymbol{q}|\boldsymbol{k}_m) &\propto p(\boldsymbol{k}_{-m}|k_{mm},\mathbf{S},\boldsymbol{q})p(\boldsymbol{q}|\boldsymbol{\alpha},\boldsymbol{\beta})p(\mathbf{S}|\boldsymbol{\pi})p(\boldsymbol{\pi}|\boldsymbol{\gamma}) \\ &= \prod_{i=1}^{N} p(k_{mi}|k_{mm},\boldsymbol{s}_i,\boldsymbol{q})p(\boldsymbol{q}|\boldsymbol{\alpha},\boldsymbol{\beta})p(\boldsymbol{s}_i|\boldsymbol{\pi})p(\boldsymbol{\pi}|\boldsymbol{\gamma}) \\ &= \prod_{i=1}^{N} {}_{k_{mm}}C_{k_{mi}} \prod_{c=1}^{C} \{\pi_c q_c^{k_{mi}}(1-q_c)^{k_{mm}-k_{mi}}\}^{s_{ic}} \\ &\times \prod_{c=1}^{C} q_c^{\alpha_c-1}(1-q_c)^{\beta_c-1}\pi_c^{\gamma_c-1}. \end{aligned}$$

さらに，$\boldsymbol{\pi}$と $\boldsymbol{q}$ を積分消去して，

$$\begin{aligned} p(\mathbf{S}|\boldsymbol{k}_m) &\propto \frac{\prod_{c=1}^{C}\Gamma(\gamma_c+\sum_{i=1}^{N}s_{ic})}{\Gamma(N+\sum_{c=1}^{C}\gamma_c)} \\ &\times \frac{\Gamma(\alpha_c+\sum_{i=1}^{N}k_{mi}s_{ic})\Gamma(\beta_c+\sum_{i=1}^{N}(k_{mm}-k_{mi})s_{ic})}{\prod_{c=1}^{C}\Gamma(\alpha_c+\beta_c+\sum_{i=1}^{N}k_{mm}s_{ic})}. \end{aligned}$$

この式は，第 m 行についての事後確率であることに注意すると，事後確率最大となる $\mathbf{S}$ は，個体 m を始点として他のすべての点への類似度のみわかる場合のクラスタリングとなる．

また，第 m 行について潜在変数とパラメータの事後確率が最大となる各潜在変数とパラメータを推定すると，パラメータ $\boldsymbol{s}_i,\ i=1,...,N,$ についての推定値は以下となる．

$$\begin{aligned} c_i &= \underset{c}{\operatorname{argmax}}\ \pi_c q_c^{k_{mi}}(1-q_c)^{k_{mm}-k_{mi}}, \\ s_{ic} &= \begin{cases} 1 & (c=c_i) \\ 0 & (\text{otherwise}). \end{cases} \end{aligned}$$

また，パラメータ $q_c,\ c=1,...,C,$ についての推定値は以下となる．

$$\begin{aligned} q_c &= \frac{\sum_{i=1}^{N}(k_{mi}s_{ic})+\alpha_c-1}{k_{mm}N_c+\alpha_c+\beta_c}, \\ N_c &= \sum_{i=1}^{N} s_{ic}. \end{aligned}$$

ここで N_c は，クラス c に所属するデータの個数を示す．さらに，π_c，$c = 1, ..., C$，の推定値は，$\sum_{c=1}^{C} \pi_c = 1$ という制約があるため，ラグランジュの未定乗数法を使い，

$$\pi_c = \frac{N_c + \gamma_c - 1}{N + C(\gamma_c - 1)}.$$

これらのパラメータは独立に最大化することができないため，初期値を与え，順に最大化することを収束するまで繰り返す．

3.2 すべての行を利用したクラスタリング

一つの行に限ってみれば，このクラスタリングで「最適な」クラスに分割できる．しかし，一つの行のみの情報しか利用しないため，クラスタリングの精度が悪い．そのため，一つの行については，対角要素の個体と同じクラスと，同じでないクラスとに 2 分する．

前節で示した一つの行に関する $\mathbf{S}$ の事後確率の最大化によるクラスの 2 分をすべての行について行う．その結果対角要素で条件づけられた事後確率 $p(\mathbf{S}|k_{ii})$ が最大となる行 m を選び，その行の 2 分割で得られたクラスタリング結果のうち個体 m と同じクラスの個体の所属クラスを Class(m) として固定する．それ以外の個体について，その個体が対角要素となる行に対して，対角要素で条件づけられた $\mathbf{S}$ の事後確率最大化により行ごとの 2 分割を行う．この手順を繰り返すことで，すべての個体の所属クラスを決定する．

3.3 K 近傍法による再クラスタリング

行ごとの事後確率最大化による 2 分割では，「局所的」な情報しか使うことができないため，「大局的」にみれば他のクラスに所属すべき個体がそうならないことが起こり得る．

そのため教師あり学習である K 近傍法を使って個体の所属クラスの更新を行う．すなわち，前節の手法によって決まった各個体の所属クラスを訓練データとして扱う．個体 i の所属クラスを更新するときは，個体 i を除くすべての個体を訓練データとし，個体 i の所属クラスを K 近傍法を用いて更新する．更新する個体はランダムに選び，個体数 N 回繰り返し更新を行う．

なお，K 近傍法とは，一般に新たな個体に対して訓練データセットから K 個の近傍にある個体の所属クラスを用いて，新たな個体の所属するクラスを推定する手法である [Cover 67]．訓練データセットに真なる所属クラスが与えられていると仮定すれば，新たな個体は類似性の高い幾つかの個体と同じ所属クラスになる確率が高いと考えられる．新たな個体の近傍にある K 個の個体の所属クラスについて多数決をとることで，新たな個体が所属するクラスが推定できる．

4. 評価実験

提案手法と既存の手法を比較することによって提案手法を評価する．既存手法として，スペクトラルクラスタリングとディリクレ過程ガウス混合モデルという中華料理店過程によるクラスタリング手法とを組み合わせた手法 (以下，SC+CRP) を用いる．本研究における評価実験では，類似度行列のみを与え，クラス数は与えない．

評価実験には，[LeCun 98] で作成された手書き数字のデータセットである MNIST を用いる．実験には，ニューラルネットワークによって次元圧縮したデータセットも合わせて利用する．MNIST とは，訓練データ 60000 個，テストデータ 10000 個の手書き数字のデータセットである．本稿では，各実験ごとに数字 1 から 4 のみを利用したデータセットと全数字を利用したデータセットの 2 種類で実験する．また，各手書き数字は 28×28 ピクセルであり，0 から 255 の整数を要素とする 784 次元のベクトルである．

評価実験に用いる類似度行列を作成するために，正規化線形カーネルを利用する．正規化線形カーネルを正規化カーネルにおいて線形カーネルを使ったカーネル関数として定義する．正規化カーネルは以下で定義される．

$$k_{norm}(\boldsymbol{x}, \boldsymbol{y}) = \frac{k(\boldsymbol{x}, \boldsymbol{y})}{\sqrt{k(\boldsymbol{x}, \boldsymbol{x})}\sqrt{k(\boldsymbol{y}, \boldsymbol{y})}}, \tag{5}$$

ここで $k(\boldsymbol{x}, \boldsymbol{y})$ は，任意のカーネル関数である．正規化線形カーネルは，$k(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ としたものである．正規化線形カーネルは，線形カーネルの値が全て 0 以上であれば，正規化線形カーネルも 0 から 1 の値をとる．そのため，正規化線形カーネルを使った類似度行列を提案手法に入力として与えるときには，全ての要素を 1000 倍し小数点以下を切り捨てることで類似度行列の各要素を整数にする．

評価尺度として，純度 (Purity)・正規化相互情報量 (NMI)・ランド指数 (RI) の 3 つを利用する．これらは [Lin 10] で使用された評価基準をであり，いずれも 0 から 1 の値をとり，1 に近いほど良い結果であるとする．さらに，クラス数を与えず実験を行うため，評価項目にクラス数を追加する．

評価実験において提案手法に以下のような設定を行う．提案手法に用いる K 近傍法による近似では，K の値を各実験で用いる個体数の 10 % とする．超パラメータは各実験ごとに設定する．

4.1 手書き数字データセットによる実験

手書き数字 MNIST のテストデータのうち 1 から 4 の数字から，それぞれランダムに 100 個ずつ抽出し，合計 400 個の個体による類似度行列 $\mathbf{K}$ に対して，SC+CRP と提案手法によってそれぞれクラスタリングを行う．SC+CRP の超パラメータは $\alpha = 0.25$ とする．また，提案手法における超パラメータを $\boldsymbol{\alpha} = (60, 581)^{\mathrm{T}}, \boldsymbol{\beta} = (76, 327)^{\mathrm{T}}, \boldsymbol{\gamma} = (362, 224)^{\mathrm{T}}$ とする．この実験環境での結果を表 1 で示す．

	SC+CRP	Proposed
Purity	0.80	0.91
NMI	0.57	0.66
RI	0.83	0.85
クラス数	6.7	12.2

表 1: MNIST の数字 1-4 から 400 個の個体を用いた評価結果

手書き数字 MNIST のテストデータの全ての数字から，それぞれランダムに 100 個ずつ抽出し，合計 1000 個の個体による類似度行列 $\mathbf{K}$ に対して，SC+CRP と提案手法を比較する．SC+CRP の超パラメータは $\alpha = 0.01$ とする．提案手法における超パラメータとしては，$\boldsymbol{\alpha} = (54, 495), \boldsymbol{\beta} = (60, 532)^{\mathrm{T}}, \boldsymbol{\gamma} = (959, 955)^{\mathrm{T}}$ とする．この実験の結果を表 2 で示す．

4.2 次元圧縮された手書き数字データセットによる実験

手書き数字データ MNIST をニューラルネットワークによって次元圧縮を行い，そのデータによる類似度行列を使い，実験を行う．手書き数字データはそれぞれ 784 次元のベクトルである．これを [Hinton 06] で提案されたニューラルネットワークによって 30 次元のベクトルに次元圧縮する．このニューラルネットワークは，制限ボルツマンマシンによって構成される

	SC+CRP	Proposed
Purity	0.60	0.53
NMI	0.53	0.49
RI	0.89	0.83
クラス数	10.0	18.8

表 2: MNIST の数字 0-9 から 1000 個の個体を用いた評価結果

3 層のネットワークである．ニューラルネットワークは，手書き数字データの訓練データ全てを用いて，自己符号化法によって学習したものである．このニューラルネットワークによってテストデータを次元圧縮し，それらに対して正規化線形カーネルを用いることで類似度行列を作成する．

まず，手書き数字データセット MNIST のテストデータのうち 1 から 4 の数字を用いて実験を行う．実験ごとにランダムに 100 個ずつ個体を選び，合計 400 個のデータセットとする．このデータセットに上記の処理を施した類似度行列 $\mathbf{K}$ に対して，比較実験を行う．SC+CRP の超パラメータは $\alpha = 0.25$ とする．提案手法における超パラメータとしては，$\boldsymbol{\alpha} = (24, 528)^{\mathrm{T}}, \boldsymbol{\beta} = (95, 809)^{\mathrm{T}}, \boldsymbol{\gamma} = (555, 279)^{\mathrm{T}}$ とする．実験の結果を表 3 を示す．

	SC+CRP	Proposed
Purity	0.76	0.89
NMI	0.52	0.62
RI	0.81	0.85
クラス数	5.8	12.7

表 3: MNIST の数字 1-4 から 400 個の個体を取り出し次元圧縮したデータセットによる評価結果

次に，全ての数字を 100 個ずつ，合計 1000 個の個体を用いて実験する．実験ごとに，ランダムにデータセット内から取り出す．同様に上記の処理を施した類似度行列 $\mathbf{K}$ に対して，比較実験を行う．SC+CRP の超パラメータは $\alpha = 0.01$ とする．提案手法における超パラメータとしては，$\boldsymbol{\alpha} = (15, 803)^{\mathrm{T}}, \boldsymbol{\beta} = (84, 122)^{\mathrm{T}}, \boldsymbol{\gamma} = (633, 567)^{\mathrm{T}}$ とする．実験の結果を表 4 に示す．

	SC+CRP	Proposed
Purity	0.60	0.54
NMI	0.54	0.46
RI	0.89	0.84
クラス数	10.0	20.3

表 4: MNIST の数字 0-9 から 1000 個の個体を取り出し次元圧縮したデータセットによる評価結果

5. 議論

実験結果より提案手法のいくつかの特性をみることができる．提案手法は，既存手法と比較して細かなクラスタリングを行う．評価実験で行ったすべての実験において，SC+CRP に比べて 2 倍程度のクラス数を示している．提案手法の特性として各行ごとに 2 分割するようなクラスタリングを行うため，小さなクラスが多く生まれ，K 近傍法による再クラスタリングを行ってもその影響が残るためと考えられる．また，クラス数以外の評価値は，SC+CRP と提案手法で大きな差はない．この結果は，提案手法がこの問題設定において，有効な手法であることを示している．さらに，類似度行列を正規化カーネルを用いて設定することで，結果が改善される．これは，行ごとの結果を比較する際，正規化されたことで行ごとの値の差が小さくなったためと考えられる．

一方，提案手法にはいくつかの課題が残る．まず，超パラメータの設定方法に関する課題である．今回の評価実験では，MNIST の訓練データを用いて適当な超パラメータを設定した．しかし，クラス数がわからないデータセットにおいて，訓練データが存在することはあまりない．様々なクラス数のデータやデータそのものの種類を変更して，多くの実験から超パラメータの特性を推定する必要がある．次に，計算時間に関する課題である．提案手法は繰り返し計算が多いため，SC+CRP に比べて計算時間がかかる．

参考文献

[Shi 00] Shi, J. and J. Malik (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888-905.

[Ng 02] Ng, A. Y., M. I. Jordan, and Y. Weiss (2002). On spectral clustering: analysis and an algorithm. *Advances in Neural Information Processing Systems*, 14, 849-856.

[Lin 10] Lin, F. and W. W. Cohen (2010). Power iteration clustering. *Proceedings of the 27th International Conference on Machine Learning* (*ICML*2010), 655-662.

[Rasmussen 00] Rasmussen, C, E. (2000). The infinite gaussian mixture model. *Advances in Neural Information Processing Systems*, 12, 554-560.

[Socher 11] Socher, R., A. L. Maas, and C. D. Manning (2011). Spectral Chinese restaurant processes: nonparametric clustering based on similarities. *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics* (*AISTATS*2011), 698-706.

[Berlinet 04] Berlinet, A. and C. Thomas-Agnan (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers.

[Cover 67] Cover, T. and P. Hart (1967). Nearest neighbor pattern classification. *IEEE Trainsactions on Information Theory* **IT-11**, 21-27.

[LeCun 98] LeCun, Y., C. Cortes, and C. J. Burges (1998). The MNIST database of handwritten digits. http://yann.lecun.com/exdb/mnist/index.html

[Hinton 06] Hinton, G. E., and R. R. Salakhutdinov (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507.

[Meila 03] Meila, M. (2003). Comparing clusterings by the variation of information. *In Learning theory and kernel machines* 173-187. Berlin, Springer.