情報論的クラスタリングに対する局所性保存グラフモデル

吉田 哲也[1,a)]

概要：本稿では，共起データを対象とする情報論的クラスタリングに対して局所性を保存するグラフモデルを提案する．近年，データ間の関係を活用する多様体学習法が注目を集めており，様々な手法が提案されている．従来の研究では，情報論的クラスタリングに対して局所的なクラスタラベルの一貫性を考慮する手法が提案されてはいるが，共起性の保存は考慮されていなかった．本稿では，近傍における相互情報量に基づいて局所性の保存を考慮するように拡張した制約付最適化問題を定式化し，拡張した問題に対するグラフモデルを提案する．本稿での提案が，同値性の観点から従来の最適化問題を含む自然な拡張であり，グラフカットに基づく近似解法の観点からも従来法の自然な拡張であることを示す．

キーワード：情報論的クラスタリング，近傍，制約付最適化，グラフカット

Locality Preserving Graph Model for Information Theoretic Clustering

Tetsuya Yoshida[1,a)]

Abstract: We propose a locality preserving graph model for information theoretic clustering. Inspired by manifold learning, previous work proposed to incorporate local consistency of cluster labels into information theoretic clustering, but local co-occurrence was not considered. We propose an extended constrained optimization problem under the framework of information theoretic clustering, and propose a locality preserving graph model for the problem. We show that the proposed model is a natural extension of previous model in terms of both the equivalence of constrained optimization and the approximate solution based on graph cut.

Keywords: information theoretic clustering, neighborhood, constrained optimization, graph cut

1. はじめに

膨大なデータを処理する場合，データ全体の性質を俯瞰するために，データ全体をいくつかのクラスタに分割するクラスタリングがしばしば利用され，ウェブ解析，情報検索などにおいても重要な役割を果たしている．

文書などのような共起データのクラスタリングに対して，データ集合とその内容を表現する集合との関係を表す結合確率に基づく情報論的クラスタリングの枠組みが提案されている [14], [16], [19]．この枠組みはクラスタリングにおけるデータ圧縮と共起の予測をともに相互情報量に基づいて制約付最適化問題として定式化するものであり，様々な近似解法が提案されてきた．他方，データ同士の関係を活用する学習法として多様体学習が提案されており [12], [17]，性能向上のための正則化にも広く活用されている [1]．

本稿では，共起データを対象とする情報論的クラスタリングに対して局所性を保存するグラフモデルを提案する．多様体学習に着想を得て，情報論的クラスタリングに対して局所的なクラスタラベルの一貫性を考慮する手法が提案されてはいるが，共起性の保存は考慮されていなかった [10]．本稿では，各データの近傍における相互情報量に基づいて局所性の保存を考慮する制約付最適化問題を定式化し，同値性の観点からこの問題が従来の最適化問題 [19] を含む自然な拡張であることを示す．次に，拡張した問題に対するグラフモデルを提案し，このモデルがグラフカットに基づく近似解法の観点から従来法 [21] の自然な拡張で

1 奈良女子大学大学院人間文化研究科
Graduate School of Humanities and Sciences, Nara Womens's University, Nara,630–8506, Japan
a) tyoshida@cc.nara-wu.ac.jp

あることを示す．

2 節で関連研究を紹介し，3 節で先行研究を説明する．4 節で提案するグラフモデルを述べ，5 節でまとめと今後の展望を述べる．

2. 関連研究

クラスタリングは大別すると階層的クラスタリングと分割的クラスタリングのアプローチがあり，従来から様々な研究が行われている [8]．前者はクラスタ階層をボトムアップに併合的に構築するアプローチとトップダウンに構築するアプローチに分けられる．他方，後者はクラスタ数などを設定して各データをクラスタに割り当てるアプローチであり，代表的な手法として k-平均法がある [5]．

クラスタリングなどの処理を行う際，処理の目的に適した表現に変換（写像）し，写像先の空間で処理を行う研究も行われている．カーネル法ではカーネルトリックを用いて高次元空間への非線形写像が用いられるが [13]，探索的データ解析などでは文書などの高次元空間で表現されるデータを低次元空間に写像して見通しのよい低次元に縮約した表現を活用することが多い．主成分分析（PCA）や多次元尺度構成法など線形写像に基づく手法 [6] に加えて，近年ではデータ対の関係に基づく多様体学習が活発に研究されている [12], [17]．たとえばスペクトルクラスタリングはデータ対の関係をグラフとして表現し，そのグラフラプラシアンから低次元の表現を学習してクラスタリングを行う手法である [20]．

性能向上や変数選択などのために正規化項の活用も広く行われており [7]，線形回帰における多重共線性に対処するためのリッジ回帰や Lasso [18] が知られている．観測データの非負性を反映して非負の表現を学習する非負値行列分解（NMF） [9] に対しても，データ対の関係に基づく多様体学習を正則化に活用する手法 [1] や直交制約を正則化に活用する手法 [4] がある．

3. 先行研究

3.1 準備

本稿では，行列は太字の大文字，ベクトルは太字のイタリック小文字で表記し，行列 $\mathbf{A}$ の第 ij 要素を a_{ij} で表す．頂点集合と辺集合から構成されるグラフにおいて，各辺に非負の重みが付いた辺重み付きグラフに対し，頂点数が n の場合に辺の重みを行列 $\mathbf{W} \in \mathbb{R}_+^{n \times n}$ で表現する [*1]．記号 $\Leftrightarrow$ は同値，$\simeq$ は近似を表す．

集合 $\mathcal{X}$ 上の確率変数 X に対する確率分布 $p_1(x)$ と $p_2(x)$ に対し，Kullback-Leibler (KL) 情報量 D_{KL} は以下で定義される [2]．

$$D_{KL}[p_1||p_2] = \sum_x p_1(x) \log \frac{p_1(x)}{p_2(x)} \tag{1}$$

$\mathcal{X}$，$\mathcal{Y}$ 上の確率変数 X，Y に対し，その結合確率を $p(x,y)$ とするとき，周辺確率をそれぞれ $p(x)$ と $p(y)$，条件付確率を $p(y|x)$ で表す．このとき，相互情報量 $I(X;Y)$ は以下で定義される．

$$I(X;Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{2}$$

$$= D_{KL}[p(x,y)||p(x)p(y)] \tag{3}$$

3.2 情報ボトルネック法

情報ボトルネック法（IB）とは共起データに対する情報論的クラスタリングの枠組み [19] であり，データ集合 $\mathcal{X}$ とその内容を表現する集合 $\mathcal{Y}$ を考え，それぞれの確率変数 X，Y に対する結合確率 $p(x,y)$ に基づいて，$\mathcal{Y}$ に対する情報を多く持つクラスタの集合 $\mathcal{T}$ を求めるものである．この枠組みでは，クラスタに対する確率変数 T は X のみに依存し Y とは無関係であるということをマルコフ関係

$$T \leftrightarrow X \leftrightarrow Y \tag{4}$$

が成り立つとして定式化し，情報論的クラスタリングを下記の制約付最適化問題として定式化した．

問題 1. 以下の目的関数 $\mathcal{L}$ を最小化する条件付確率 $p(t|x)$ を求めよ．

$$\mathcal{L}[p(t|x)] = I(X;T) - \beta I(T;Y) \tag{5}$$

$I(X;T)$ と $I(T;Y)$ はそれぞれ X と T，T と Y の相互情報量であり，β はパラメータである．

データ集合 $\mathcal{X}$ をクラスタ集合 $\mathcal{T}$ として圧縮して表現した際，$\mathcal{Y}$ について情報を多く持つ $\mathcal{T}$ を求めるという問題設定において，圧縮の程度を $I(X;T)$，予測の程度を $I(T;Y)$ で表現している．

問題 1 に対する最適解は以下の条件を満たす必要があることが知られている [19]．

定理 1. 結合確率 $p(x,y)$ と β が与えられ，マルコフ関係（*4*）が成立する場合，条件付き確率 $p(t|x)$ は以下の式を満たす場合に限り $\mathcal{L}$ の定常分布となる．

$$p(t|x) = \frac{p(t)}{Z(x,\beta)} exp(-\beta D_{KL}[p(y|x)||p(y|t)]) \tag{6}$$

$$Z(x,\beta) = \sum_t p(t) exp(-\beta D_{KL}[p(y|x)||p(y|t)]) \tag{7}$$

3.3 近似解法

問題 1 に対する大域的最適解を求めることは困難であるため，逐次射影や階層的併合，クラスタ分岐などに基づく近似解法が提案されてきた [15], [19]．その中で sIB と呼ぶ手法が他手法より優れていることが報告されている [16]．sIB とは，問題 1 と同値な最適化問題である

[*1] 第 ij 要素 w_{ij} は頂点対 (v_i, v_j) の辺の重みを表し，辺がない場合には重みは 0 とする．

$$\mathcal{L}_{max}[p(t|x)] = I(T;Y) - \beta^{-1} I(X;T) \quad (8)$$

の最大化を各データのクラスタへの逐次再割り当てにより近似的に解く手法である．

我々も，定理 1 と式 (6) に基づき，

$$s(x_i, x_j) = p(x_j) exp(-\beta D_{KL}[p(y|x_i)||p(y|x_j)]) \quad (9)$$

と定義するデータ間の類似度 $s(\cdot,\cdot)$ を提案し，この類似度を重み w_{ij} としてデータ集合を辺重み付きグラフとして表現するグラフモデルを提案し，このグラフの最小カットを求める問題が問題 1 の近似問題となることを示した [21]．さらに，このグラフモデルの最小カットを求める際，酔歩に基づく正規化グラフラプラシアンを用いるスペクトルクラスタリング [20] を適用し，その有効性を報告した．本稿ではこの手法を gIB と呼ぶ．

3.4 局所的な一貫性を考慮する sLCIB

多様体学習に着想を得て，局所的なクラスタラベルの一貫性を考慮して sIB を拡張した sLCIB が提案された [10]．sLCIB では，まず式 (8) における β^{-1} を 0 に限定して第 2 項を無視し，次に各データ x_k の近傍 N_k における相互情報量 $I_{N_k}(X;T)$ を局所的なクラスタラベルの一貫性を表す正則化項として追加した目的関数

$$\mathcal{L}_{max}^{LC_{org}}[p(t|x)] = I(T;Y) - \mu \sum_{x_k \in X} I_{N_k}(X;T) \quad (10)$$

の最大化を考える．式（10）で μ は正則化パラメータであり，近傍 N_k の範囲もパラメータとなる．

さらに，$p(t|x)$ を 0 か 1 に限定 *2 して $\mathcal{L}_{max}^{LC_{org}}[p(t|x)]$ を近似し，各データ x_k の $I_{N_i}(X;T)$ を近傍 N_k におけるクラスタラベルの情報エントロピー $H_{N_k}(T)$ とした

$$\mathcal{L}_{max}^{LC}[p(t|x)] = I(T;Y) - \mu \sum_{x_k \in X} H_{N_k}(T) \quad (11)$$

の最大化を，sIB と同様にデータの逐次再割り当てを通じて近似的に行う．

4. 局所性を保存するグラフモデル

4.1 先行研究での課題

3.4 節で述べた先行研究では近傍に基づく正則化項の活用を提案しているが，以下の課題がある．

i) IB は共起データの結合確率に基づく枠組みであるが，近傍における結合確率から定式化されていない．

ii) データ全体における条件付き確率と近傍における条件付き確率の同時最適化が考慮されていない．

iii) 式（5）では $I(X;T)$ と $I(T;Y)$ の両者を考慮するが，近傍では $I(X;T)$ しか考慮されていない．

上記の課題に対し，本稿では 4.2 節でまず近傍における結合確率を定義し，この定義から近傍における様々な確率分布や相互情報量を導出する．次に，局所性の保存を考慮する制約付最適化問題を定式化し，この問題が制約付最適化問題の同値性の観点から問題 1 の自然な拡張であることを示す．さらに，拡張した問題に対するグラフモデルを提案し，このモデルが，グラフカットに基づく近似解法の観点から従来法 [21] の自然な拡張であることを示す．

*2 ハードクラスタリングに限定することを意味する．

4.2 近傍における確率分布

IB の枠組みではデータ全体での結合確率 $p(x,y)$ は与えられると仮定するため，データ x_k の近傍 N_k における結合確率 $p_{N_k}(x,y)$ を条件付き確率として以下で定義する．

$$p_{N_k}(x,y) = \frac{p(x,y)}{\sum_{x \in N_k} \sum_y p(x,y)} \quad (12)$$

ただし $N_k = \phi$ の場合は $p_{N_k}(x,y) = 0$ と定義し，以下で述べる近傍での相互情報量も 0 となる．

$p_{N_k}(x,y)$ の周辺確率として近傍における確率 $p_{N_k}(x)$，$p_{N_k}(y)$ をそれぞれ定義し，この周辺確率から条件付き確率 $p_{N_k}(y|x)$ を定義すると，以下が成り立つ．

$$p_{N_k}(x) = \frac{p(x)}{\sum_{x \in N_k} p(x)} \quad (13)$$

$$p_{N_k}(y|x) = p(y|x) \quad (14)$$

証明は割愛する．式（13）は，近傍における結合確率から導出した $p_{N_k}(x)$ もまた近傍における条件付き確率と解釈できることを意味し，式（14）はデータ全体と近傍とで条件付き確率が不変であることを意味する．

次に，上記 ii) の同時最適化の観点から，求める条件付き確率は以下を満たすものとする．

$$p_{N_k}(t|x) = p(t|x) \quad (15)$$

式（15）と式（4）のもとで，近傍でのクラスタの確率 $p_{N_k}(t)$ は以下となる．

$$p_{N_k}(t) = \sum_{x \in N_k} \sum_y p_{N_k}(x,y) p(t|x) \quad (16)$$

上記に基づき，近傍における相互情報量は下記となる．

$$I_{N_k}(X,T) = \sum_{x \in N_k} p_{N_k}(x) \sum_t p(t|x) \log \frac{p(t|x)}{p_{N_k}(t)} \quad (17)$$

$$I_{N_k}(T,Y) = \sum_t \sum_y p_{N_k}(t,y) \log \frac{p_{N_k}(t,y)}{p_{N_k}(t) p_{N_k}(y)} \quad (18)$$

4.3 局所性の保存を考慮した制約付最適化問題

IB における問題 1 を拡張し，局所性の保存を考慮して近傍における相互情報量を正則化項とする以下の制約付最適化問題を考える．

問題 2. 以下の目的関数 $\mathcal{L}^{LP}$ を最小化する条件付確率 $p(t|x)$ を求めよ．

$$\mathcal{L}^{LP}[p(t|x)] = (I(X;T) - \beta I(T;Y)) + \mu \sum_{x_k \in X} (I_{N_k}(X;T) - \beta I_{N_k}(T;Y)) \quad (19)$$

β は式（5）のパラメータ，μ は正則化係数，N_k はデータ x_k の近傍である．

先行研究での式（10），（11）における上記 **iii)** の課題に対して，式（19）では IB の枠組みに沿って両方の相互情報量を考慮した目的関数となっている．

位相空間では様々な近傍が考えられるが [11]，極端な位相として，離散位相（$N_k = \phi$）の場合に式（19）の正則化項は 0 となり問題 1 と一致する．逆に，密着位相（$N_k =\mathcal{X}$）の場合には $\mathcal{L}^{LP}[p(t|x)]$ は $\mathcal{L}[p(t|x)]$ の定数倍となるため問題 1 と問題 2 は同値な問題となる．他方，先行研究での式（10）と（11）は両方の位相に対してさえも同値な問題とはなっていない．この意味で，提案する問題 2 は問題 1 を含む自然な拡張であると考えられる．

次に，近傍での式（7）に対応する関数を以下で定義する．

$$Z_{N_k}(x,\beta) = \sum_t p_{N_k}(t) exp(-\beta D_{KL}[p(y|x)||p_{N_k}(y|t)]) \quad (20)$$

式（7），（20）に基づく以下の最小化問題を考える．

問題 3. 以下の目的関数を最小化する $p(t|x)$ を求めよ．

$$F^{LP} = \sum_x \sum_t p(x)p(t|x)(-\log Z(x,\beta)) + \mu \sum_{x_k \in X} \sum_{x \in N_k} \sum_t p(x)_{N_k} p(t|x)(-\log Z_{N_k}(x,\beta)) \quad (21)$$

定理 2. 問題 *2* と問題 *3* は同値な問題である．

証明は割愛する．F^{LP} は近傍を考慮する場合の自由エネルギーである．もとの IB の枠組みにおいて式（19）の第 1 項の最小化と式（21）の第 1 項の最小化が同値であったが，正則化項として第 2 項を加えて拡張した場合にも同様に同値性が成り立つことを示しており，4.2 節で述べた近傍における結合確率に基づく拡張の妥当性を示していると考えられる．

4.4 局所性を保存するグラフモデル

IB における問題 1 は変数 $p(t|x)$ に対して線形なため定理 1 の定常分布の条件を導出できるが，近傍を考慮して拡張した問題 2 は，近傍における確率分布が式（12）や（13）のように条件付き確率となって分母にも $p(t|x)$ が影響するため非線形となり，定常分布の条件を導出することは困難である．そこで，定理 2 に基づいて問題 3 を考え，式（21）の F^{LP} の近似問題を解くことを考える．

データ対 (x_i, x_j) の間の辺の重み w_{ij} を式（9）を用いて定義し，データ集合 $\mathcal{X}$ を重み付きグラフとして表現し，データ間の遷移確率を

$$p(x_j|x_i) = \frac{w_{ij}}{\sum_j w_{ij}} \quad (22)$$

と定義すると，この条件付き確率はクラスタ集合 $\mathcal{T}$ を $\mathcal{X}$ とした場合 [*3] の定理 1 の定常分布となり問題 1 の最適解であるため，グラフの分割による最適解からの乖離の最小化をグラフカットにより近似的に解くことが提案された [21]．そこで，式（9）に基づいてまずデータ集合 $\mathcal{X}$ を重み付きグラフとして表現し，このグラフにデータ x_k ごとに近傍 N_k を考慮した重み付きグラフを重ね合わせたグラフを考え，そのグラフに対するカットにより F^{LP} の近似問題を解くことを考える．

データ対 (x_i, x_j) に対し，データ集合 $\mathcal{X}$ における上記の重み w_{ij} と，以下で定義するデータ x_k の近傍 N_k を反映する重み w_{ij}^k を考える．

$$w_{ij} = p(x_j) exp(-\beta D_{KL}[p(y|x_i)||p(y|x_j)]) \quad (23)$$

$$w_{ij}^k = I_{ij}^k p_{N_k}(x_j) exp(-\beta D_{KL}[p(y|x_i)||p(y|x_j)]) \quad (24)$$

$$I_{ij}^k = \begin{cases} 1 & x_i, x_j \in N_k \\ 0 & \text{otherwise} \end{cases} \quad (25)$$

上記の重みを持つグラフ G，G^k を考え，各グラフ上での遷移確率 $p(x_j|x_i)$，$p_{N_k}(x_j|x_i)$ を式（22）と同様に定義し，x_i に接続する辺の重みの総和を以下で表す．

$$d_i = \sum_{x_j \in X} w_{ij}, \quad d_i^k = \sum_{x_j \in X} w_{ij}^k \quad (26)$$

グラフの重ね合わせとして，局所性を保存するグラフモデルとして以下の重み付きグラフ G^{LP} を定義する．

$$w_{ij}^{LP} = w_{ij} + \mu \sum_{x_k \in X} w_{ij}^k \quad (27)$$

式（13）を代入して整理すると以下となる．

$$w_{ij}^{LP} = \left(1 + \mu \sum_{x_k \in X} I_{ij}^k \frac{1}{\sum_{x \in N_k} p(x)}\right) w_{ij} \quad (28)$$

式（28）はもとのグラフモデル G における重み w_{ij} に近傍の影響を表す係数が掛けられたものであることを示す．

4.5 グラフカットによる近似解法

以下ではデータが一様分布すると仮定し，$p(x) = c$（$c > 0$ の定数）とする．このとき，グラフ G，G^k の自由エネルギー F_G，F_G^k はそれぞれ c に比例する定数となる（[21] 命題 4）．最適化問題では定数係数 c を無視しても同値な問題であるため，以降の議論では係数を省いて記述する．F^{LP} に対応する自由エネルギーとして，G^{LP} に対する $F_G^{LP} = F_G + \mu \sum_{x_k \in X} F_G^k$ を考え，データ集合 $\mathcal{X}$ を $\mathcal{X} = S \sqcup \bar{S}$ と分割した際，分割前の F_G^{LP} からの乖離が最小となる分割を求める問題を考える．ここで，分割前の自由エネルギーはそれぞれ下記となる．

[*3] 階層的クラスタリングにおける初期状態で各データをクラスタとみなすことに対応する．

$$F_G = \sum_{x_i \in X} (-\log d_i) \tag{29}$$

$$F_G^k = \sum_{x_i \in N_k} (-\log d_i^k) \tag{30}$$

$$\begin{aligned} F_G^{LP} &= F_G + \mu \sum_{x_k \in X} F_G^k \\ &= \sum_{x_i \in X} (-\log d_i) + \mu \sum_{x_k \in X} \sum_{x_i \in N_k} (-\log d_i^k) \end{aligned} \tag{31}$$

データ集合の分割 $\mathcal{X} = S \sqcup \bar{S}$ によりグラフ G^{LP} は誘導部分グラフ G_S^{LP} と $G_{\bar{S}}^{LP}$ に分割され，分割後のグラフは非連結なグラフ $\hat{G}^{LP} = \{G_S^{LP}, G_{\bar{S}}^{LP}\}$ となる [3].

各データ x_i に対し，$\mathcal{X}$ において x_i を含む部分集合を S_i，補集合を $\bar{S}_i$ と表し，式（26）に準じて以下を定義する．

$$d_i^S = \sum_{x_j \in S_i} w_{ij}, \quad d_i^{\bar{S}} = \sum_{x_j \in \bar{S}_i} w_{ij} \tag{32}$$

$$d_i^{kS} = \sum_{x_j \in S_i} w_{ij}^k, \quad d_i^{k\bar{S}} = \sum_{x_j \in \bar{S}_i} w_{ij}^k \tag{33}$$

分割後の G_S^{LP}，$G_{\bar{S}}^{LP}$ に対する自由エネルギー $F_{G_S}^{LP}$，$F_{G_{\bar{S}}}^{LP}$ は式（31）と同様に下記となり，分割後のグラフ $\hat{G}^{LP}$ の $F_{\hat{G}}^{LP}$ は両者の和となる．位相空間での部分空間に対する近傍の定義 [11] より，近傍 N_k は分割により部分空間 S，$\bar{S}$ において以下の N_k^S，$N_k^{\bar{S}}$ となることに注意されたい．

$$N_k^S = N_k \cap S, \quad N_k^{\bar{S}} = N_k \cap \bar{S} \tag{34}$$

$$F_{G_S}^{LP} = \sum_{x_i \in S} (-\log d_i^S) + \mu \sum_{x_k \in X} \sum_{x_i \in N_k^S} (-\log d_i^{kS}) \tag{35}$$

$$F_{G_{\bar{S}}}^{LP} = \sum_{x_i \in \bar{S}} (-\log d_i^{\bar{S}}) + \mu \sum_{x_k \in X} \sum_{x_i \in N_k^{\bar{S}}} (-\log d_i^{k\bar{S}}) \tag{36}$$

分割前の F_G^{LP} は定数であるため，分割による乖離 $(F_{G_S}^{LP} + F_{G_{\bar{S}}}^{LP}) - F_G^{LP}$ の最小化は $(F_{G_S}^{LP} + F_{G_{\bar{S}}}^{LP})$ の最小化と同値な問題となる．G^{LP} の分割におけるグラフカットを以下で定義する．

$$cut(S, \bar{S}) = \sum_{x_i \in S} \sum_{x_j \in \bar{S}} w_{ij}^{LP}, \; cut(\bar{S}, S) = \sum_{x_i \in \bar{S}} \sum_{x_j \in S} w_{ij}^{LP} \tag{37}$$

このとき，近傍を考慮しない場合 [21] と同様に以下が成り立つ．

主張 3. F^{LP} の最小化は $cut(S, \bar{S}) + cut(\bar{S}, S)$ の最小化として近似できる．

証明は割愛する．主張 3 より，問題 2 を局所性を保存するグラフモデルに対して最小カットを求める問題に近似できる．クラスタリングを行う際には，データ集合を提案するグラフモデルで表現し，gIB [21] と同様に酔歩に基づく正規化グラフラプラシアンを用いるスペクトルクラスタリングを用いる．この手法を gLPIB と呼ぶ．

5. おわりに

本稿では，共起データを対象とする情報論的クラスタリングに対して局所性を保存するグラフモデルを提案した．局所的なクラスタラベルの一貫性に基づく従来法の課題を指摘するとともに，近傍における相互情報量に基づいて局所性の保存を考慮する制約付最適化問題を定式化し，同値性の観点から従来の最適化問題を含む自然な拡張であることを示した．さらに，拡張した問題に対するグラフモデルを提案し，このモデルがグラフカットに基づく近似解法の観点から従来法の自然な拡張であることを示した．特に，位相空間における部分空間の観点から分割後のグラフモデルにおける近傍を考えることでグラフカットとの対応付けが可能となることを示した．

謝辞　本研究の一部は文部科学省科研費 (No. 15K00307) の補助による．

参考文献

[1] Cai, D., He, X., Wu, X. and Han, J.: Non-negative Matrix Factorization on Manifold, *Proc. ICDM'08*, pp. 63–72 (2008).
[2] Cover, T. and Thomas, J.: *Elements of Information Theory*, Wiley (2006).
[3] Diestel, R.: *Graph Theory*, Springer (2006).
[4] Ding, C., Li, T., Peng, W. and Park, H.: Orthogonal Nonnegative Matrix Tri-factorizations for Clustering, *Proc. KDD '06*, pp. 126–135 (2006).
[5] Hartigan, J. and Wong, M.: Algorithm AS136: A k-means clustering algorithm, *Journal of Applied Statistics*, Vol. 28, pp. 100–108 (1979).
[6] Harville, D. A.: *Matrix Algebra From a Statistican's Perspective*, Splinger (2008).
[7] Hastie, T., Tibshirani, R. and Friedman, J.: *The Elements of Statistical Learning*, Springer (2008).
[8] Jain, A., Murty, M. and P.J., F.: Data Clustering: A Review, *ACM Computing Surveys*, Vol. 31, pp. 264–323 (1999).
[9] Lee, D. D. and Seung, H. S.: Algorithms for Non-negative Matrix Factorization, *Proc. NIPS'01*, pp. 556–562 (2001).
[10] Lou, Z., Ye, Y. and Zhu, Z.: Information Bottleneck with local consistency, *Proc. PRICAI'12*, pp. 285–296 (2012).
[11] 松阪和夫：集合・位相入門，岩波書店 (1968).
[12] Roweis, S. and Saul, L.: Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science*, Vol. 290, No. 22, pp. 2323–2326 (2000).
[13] Shawe-Taylor, J. and Cristianini, N.: *Kernel Methods for Pattern Analysis*, Cambridge University Press (2004).
[14] Slonim, N. and Tishby, N.: Agglomerative Information Bottleneck, *Proc. NIPS'00*, pp. 617–623 (2000).
[15] Slonim, N.: The Information Bottleneck:Theory and Applications, PhD Thesis, Hebrew University (2002).
[16] Slonim, N., Friedman, N. and Tishby, N.: Unsupervised Document Classification using Sequential Information Maximization, *SIGIR-02*, pp. 129–136 (2002).
[17] Tenenbaum, J., de Silva, J. and Langford, J.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science*, Vol. 290, No. 22, pp. 2319–2323 (2000).
[18] Tibshirani, R.: Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Series B*, Vol. 58, No. 1, pp. 267–288 (1996).

[19] Tishby, N., Pereira, F. and Bialek, W.: The Information Bottleneck Method, *Proc. of the 37th Allerton Conference on Communication and Computation*, pp. 368–377 (1999).

[20] von Luxburg, U.: A Tutorial on Spectral Clustering, *Statistics and Computing*, Vol. 17, No. 4, pp. 395–416 (2007).

[21] 吉田哲也：相互情報量に基づくクラスタリングに対するグラフモデルとその評価，情報処理学会論文誌:数理モデル化と応用，Vol. 3, No. 3, pp. 1–10 (2010).