情報処理学会論文誌　数理モデル化と応用　Vol. 3　No. 3　1–11 (Oct. 2010)

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

吉　田　哲　也[†1]

本稿では，相互情報量に基づくクラスタリング問題に対するグラフモデルを提案する．相互情報量から導出される定常分布に着想を得たデータ間の類似度関数を定義してデータ集合を辺重み付きグラフとして表現することにより，データが一様分布する場合にはハードクラスタリング問題が提案するグラフモデルにおける組合せ最適化問題に近似できることを示す．グラフモデルを用いてデータ集合を表現することにより，グラフ構造に基づく様々なアルゴリズムを用いてクラスタリングを行うことが可能になる．提案するグラフモデルを文書クラスタリングでのベンチマークデータである 20 Newsgroup のデータに対して評価し，他手法との比較を通じて提案手法の妥当性と有効性を確認した．

A Graph Model for Mutual Information Based Clustering and Its Evaluation

Tetsuya Yoshida[†1]

We propose a graph model for data clustering based on mutual information. Based on the stationary distribution induced from the problem setting, we propose a similarity function among data objects, and represent the entire objects as an edge-weighted graph. We show that, in hard assignment, the problem can be approximated as a combinatorial problem over the proposed graph when data is uniformly distributed. Various algorithms can be utilized to solve the clustering problem by representing the dataset based on our graph model. The proposed approach is evaluated on the text clustering problem over the 20 Newsgroup benchmark data. The results are encouraging and indicate the effectiveness of our approach.

1. はじめに

近年の高速なネットワークと廉価な大容量記憶装置の普及により，科学，技術，社会に関するきわめて多くの情報が電子的に可読なデータとして公開され，広く利用されている．アクセス可能なデータ量の増加にともない，膨大なデータからの有用な知識発見を目指すデータマイニングの手法が鋭意研究開発されている．たとえばバスケット分析に代表される頻出パターンの抽出や分類ルールの学習などに対して多くの研究がなされ，様々な分野で多大な成果をあげている．しかし，膨大なデータを扱う場合，個々のデータを対象とした処理に取り組む前に，扱うべきデータ全体の性質を把握することが役立つことも多い．データ全体の性質を俯瞰するために，データ間の関係（類似性など）に基づいて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手法がしばしば利用される．

クラスタリングとは，類似するデータは同じグループに割り当てられ，類似しないデータは異なるグループに割り当てられるように，データ全体をいくつかのグループ（クラスタ）に分割する処理である．クラスタリングは，膨大なデータを対象とするウェブ解析，情報検索などにおいても重要な役割を果たしている．

本稿では，相互情報量に基づくクラスタリングの枠組み[23] について考察する．この枠組みでは，データクラスタリングは相互情報量に基づく制約付最適化問題として定式化されるが，相互情報量の非線形性や目的関数の非凸性などにより大域的最適解を得ることが困難なため，様々な近似解法が提案されてきた[18),20),23)]．

本稿では，相互情報量に基づくクラスタリング問題に対するグラフモデルを提案し，データが一様分布する場合にはハードクラスタリング問題が提案するグラフモデルにおける組合せ最適化問題に近似できることを示す．まず，相互情報量から導出される定常分布に着想を得たデータ間の類似度関数を定義し，この類似度関数に基づいてデータ集合全体を辺重み付きグラフとして表現することを提案する．各データをグラフの頂点に対応させ，頂点対を類似度関数から得られる類似度を重みとする辺で連結することにより，データ集合を辺重み付きグラフとして表現する．次に，相互情報量に基づくクラスタリング問題が，提案するグラフモデルによりデータ集合を表現したグラフ上の組合せ最適化問題として定式化できることを示す．

†1 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University

データ集合を本稿で提案するグラフモデルに基づいて辺重み付きグラフとして表現することにより，グラフ構造に基づく様々なアルゴリズムを用いてクラスタリングを行うことが可能になる．たとえば，グラフの最小カット問題に対応するスペクトルクラスタリング法[24]を用いて文献 23) での制約付最適化問題を解くことが可能となる．提案手法を文書クラスタリングでのベンチマークデータである 20 Newsgroup のデータに対して評価し，他手法との比較を通じて提案手法の妥当性と有効性を確認した．特に，クラスタ数が多く困難な問題に対する提案手法の有効性を確認した．

2 章では関連研究を紹介し，3 章では相互情報量に基づくクラスタリングの枠組みの概略を説明する．4 章で提案するグラフモデルの詳細を説明し，3 章の問題がグラフモデル上の問題として定式化できることを示す．5 章で提案手法の評価を報告し，6 章で本稿での提案に関する議論を述べた後，7 章でまとめと今後の展望について述べる．

2. 関 連 研 究

クラスタリングには，大きく分けて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る[12]．前者はデンドログラムと呼ばれる木構造としてクラスタ階層を構築してデータを分割する．木構造における葉ノードには各データが割り当てられ，1 つのデータのみからなるクラスタを表現する．内部ノードはその子ノードに対応するクラスタ集合を併合したクラスタに対応し，入れ子になったクラスタを表現する．さらに，階層的クラスタリング手法はクラスタ階層をボトムアップに構築する併合的なアプローチとトップダウンに構築する分割的なアプローチに分けられる．

分割的クラスタリングでは，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てデータ分割を生成する．クラスタの表現としては，k-平均法のように[10] セントロイドやセントロイド近隣のデータを用いた表現[15] が利用される．各データのクラスタへの割当て規範としては，ユークリッド距離などの相互の距離に基づく手法[10] や，データの生成モデルを仮定して最尤法を用いる手法[3]，データの確率分布に基づく手法[5),16] などが提案されている．

分割的クラスタリングにおいて，データ集合をグラフ構造の観点からとらえて，グラフ理論に基づいてクラスタリングを行うアプローチがある．このアプローチでは，類似度（あるいは非類似度）に基づいてグラフの中から組合せ論的な構造を探索する．これまで，グラフの彩色に基づく手法[7),9] やグラフのカットに基づく手法[24] などが提案されている．本稿でのアプローチはグラフ構造に基づく分割的クラスタリングに位置づけられる．

3. 問 題 設 定

3.1 準　　備

以下では，$\boldsymbol{X}$ で（与えられた）データ集合を表現し，$|\boldsymbol{X}|$ で集合の大きさ（要素数）を表現する．$\mathcal{X}$ 上の確率変数 X に対する確率分布 $p_1(x)$ と $p_2(x)$ を考える．

定義 1. $\mathcal{X}$ 上の確率変数 X に対する確率分布 $p_1(x)$ と $p_2(x)$ に対する *Kullback-Leibler*（*KL*）情報量は以下で定義される[2]．

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

$\mathcal{X}$，$\mathcal{Y}$ 上の確率変数 X，Y に対し，結合確率分布を $p(x,y)$ と表記し，$p(x)$ と $p(y)$ をその周辺確率分布とする．また，条件付き確率分布を $p(y|x)$ と表記する．

定義 2. *2* つの確率変数 X と Y の間の相互情報量 $I(X;Y)$ は以下で定義される．

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

$$= \sum_x p(x) \sum_y p(y|x) \log \frac{p(y|x)}{p(y)} \tag{3}$$

$$= D_{KL}[p(x,y)||p(x)p(y)] \tag{4}$$

3.2 情報ボトルネック法

相互情報量に基づくクラスタリングの枠組みは文献 23) で情報ボトルネック法として提案された．この手法は文献 16) の手法と同様，データ集合 $\boldsymbol{X}$ の確率分布に基づいてクラスタリングを行うが，与えられたデータ集合に対する関連変数 Y を導入し，Y に対する情報を多く持つようなクラスタの集合 $\boldsymbol{T}$ を求める問題ととらえる枠組みである．

この枠組みにおいては，確率変数 T は X のみに依存して Y には無関係である，ということが以下のマルコフ関係として定式化される．

定義 3. 確率変数 X，Y，T に対して以下のマルコフ関係が成り立つ．

$$T \leftrightarrow X \leftrightarrow Y \tag{5}$$

一例として，文書集合 $\boldsymbol{X} = \{x_1, \ldots, x_n\}$ が与えられ，各文書 x_i は文書中の単語の頻度ベクトルとして表現される場合を考える．この場合，文書集合 $\boldsymbol{X}$ を表現するのに使用された単語の集合が $\boldsymbol{Y} = \{y_1, \ldots, y_m\}$ に対応する．$p(x,y)$ は文書 x と単語 y の同時確率に対応し，たとえば文書 x と単語 y の共起回数に基づいて推定される．クラスタリングの目的

3　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

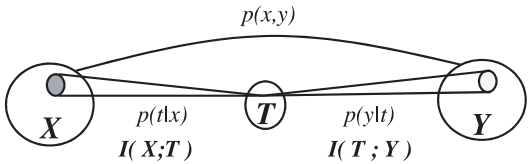

図 1　相互情報量に基づくクラスタリングの枠組み[23)]
Fig. 1　Data clustering based on mutual information [23)].

は，各クラスタ t が単語の予測に有効であるようなデータ集合 $\boldsymbol{X}$ に対するクラスタ集合 $\boldsymbol{T} = \{t_1, \ldots, t_k\}$ を見つけることである．

定義 3 に基づき，文献 23) は相互情報量に基づくクラスタリングを制約付き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き確率 $p(t|x)$ を求めよ．

$$\mathcal{L} = I(X;T) - \beta I(T;Y) \tag{6}$$

$I(X;T)$ 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 はハイパーパラメータである．

問題 1 での枠組みを**図 1** に示す．直観的には，データ集合 $\boldsymbol{X}$ をクラスタ集合 $\boldsymbol{T}$ に圧縮して表現し，圧縮した表現 $\boldsymbol{T}$ が $\boldsymbol{Y}$ について情報を多く持つ（予測できる）という問題設定を，$\boldsymbol{X}$，$\boldsymbol{T}$，$\boldsymbol{Y}$ に対応する確率変数 X，T，Y を考え，圧縮の程度を相互情報量 $I(X;T)$ で表現し，また予測の程度を相互情報量 $I(T;Y)$ で表現することにより，両者を相互情報量に基づくクラスタリング問題として定式化している[⋆1]．

問題 1 に対する最適解は以下に示す式を満たす必要がある[23)]．

定理 1. 同時確率 $p(x,y)$ と β が与えられ，マルコフ関係 (5) が成立する場合には，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．

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

$$Z(x,\beta) = \sum_t p(t) exp(-\beta D_{KL}[p(y|x)||p(y|t)]) \tag{8}$$

⋆1 $\mathcal{L}$ の最小化問題のため，$I(X;T)$ の最小化と $-I(T;Y)$ の最小化（$I(T;Y)$ の最大化）を考えることになる．前者はできるだけ圧縮した表現を生成することに対応し，後者はできるだけ予測の精度が良い表現を生成することに対応する．

3.3　従来の近似解法

定理 1 は問題 1 に対する解である条件付き確率 $p(t|x)$ が前節の問題設定における定常分布であることを示すが，式 (7) における左辺の $p(t|x)$ は同時に右辺にも（非線形に）影響をおよぼす．さらに，式 (6) での目的関数 $\mathcal{L}$ は $p(t|x)$，$p(t)$，$p(y|t)$ に対して同時には凸ではない．このため，問題 1 に対する大域的最適解を求めることは非常に困難な問題となる．

上記の問題に対し，これまで近似解を求めるいくつかの解法が提案されてきた[18),20),23)]．その中で，sIB 法と呼ばれる手法が計算量や得られるクラスタの質の観点から他の手法よりも良いことが報告されている[20)]．この手法は，問題 1 に双対な最大化問題をデータの逐次再割当てによって近似的に解く手法であり，各データを単一のクラスタに割り当てるという意味でハードクラスタリングを行う．

4.　提案するグラフモデル

4.1　準　　備

頂点集合 $\boldsymbol{V}$ と辺集合 $\boldsymbol{E} \in \boldsymbol{V} \times \boldsymbol{V}$ から構成されるグラフを $G(\boldsymbol{V}, \boldsymbol{E})$ と表記する．辺重み付きグラフ $G(\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{W})$ は各辺に重みが付いたグラフであり，$\boldsymbol{W}$ は重みの集合である．$|\boldsymbol{V}| = n$ の場合，重みの集合は $n \times n$ 行列 $\mathbf{W}$ で表現することができ[⋆2]，行列 $\mathbf{W}$ の第 ij 要素は頂点対 (v_i, v_j) の間の辺に対する重みを表す．なお，辺がない頂点対間での重みは 0 とする．

4.2　データ間の類似度関数

本稿では，定理 1 と式 (7) に基づき，データ x とクラスタ t の間の KL 情報量 $D_{KL}[p(y|x)||p(y|t)]$ が 3 章での枠組みにおける非類似度を表現するととらえる．さらに，この非類似度を $\mathcal{X} \times \mathcal{T}$ から $\mathcal{X} \times \mathcal{X}$ に拡張して，KL 情報量をデータ間の非類似度を表現する関数ととらえる．

上記の非類似度関数に基づき，3 章の枠組みにおけるデータ間の類似度関数として以下を提案する．

定義 4. $s : \mathcal{X} \times \mathcal{X} \to \mathcal{R}^+$ は以下で定義されるデータ間の類似度関数である．

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

β は問題 *1* でのハイパーパラメータである．

⋆2 太字のイタリック文字は集合，太字は行列の表記に対応する．

4.3　データグラフ

式 (9) で定義した類似度関数は，データ集合 $\boldsymbol{X}$ での任意のデータ対 (x_i, x_j) の間の関係を類似度として表現する．データ対の関係はグラフとして表現できるため，3 章の問題設定においてデータ集合 $\boldsymbol{X}$ を式 (9) で計算される類似度を重みとする辺重み付きグラフとして表現するグラフモデルを提案する．

定義 5. データ集合 $\boldsymbol{X}$ に対する辺重み付きグラフ $G(\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{W})$ を以下で定義する．

$$\boldsymbol{V} = \boldsymbol{X} \tag{10}$$

$$w_{ij} = \begin{cases} s(x_i, x_j) & x_i \neq x_j \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$\boldsymbol{E} = \{(x_i, x_j) | w_{ij} > 0\} \tag{12}$$

式 (10) より個々のデータと頂点は 1 対 1 に対応するため，以下では $\boldsymbol{X}$ でデータグラフにおける頂点集合も表記することとする．定義 4 よりすべての重みは非負（$w_{ij} \geq 0$, $\forall x_i, x_j \in \boldsymbol{X}$）である．本稿では，上記で定義したグラフを**データグラフ**と呼ぶ．以下では，データグラフ G は連結グラフであると仮定する[⋆1]．

命題 2. $\frac{w_{ij}}{\sum_j w_{ij}}$ はデータグラフにおいて頂点 x_j の頂点 x_i に対する条件付き確率である．

証明　定義 4 と式 (11) より w_{ij} は非負であり，G は連結グラフのため $\sum_j w_{ij} > 0$ である．$0 \leq \frac{w_{ij}}{\sum_j w_{ij}} \leq 1$ が成立し，また $\sum_j \frac{w_{ij}}{\sum_j w_{ij}} = 1$ である．このため $\frac{w_{ij}}{\sum_j w_{ij}}$ は確率の公理を満たす．□

命題 2 より，データグラフにおける条件付き確率を以下で定義する．

$$p(x_j | x_i) = \frac{w_{ij}}{\sum_j w_{ij}} \tag{13}$$

式 (13) の条件付き確率はデータグラフ上でのデータ x_i からデータ x_j への遷移確率と解釈できる．

命題 3. 式 (13) の条件付き確率は $\boldsymbol{T} = \boldsymbol{X}$ の場合における定理 *1* の定常分布である．

証明　各データ x_j をクラスタ t と見なすことにより，式 (7)，(8)，(9) から確認できる．□

データ集合 $\boldsymbol{X}$ をクラスタ集合 $\boldsymbol{T}$ と見なして $\boldsymbol{T} = \boldsymbol{X}$ とすることは，2 章で紹介した階層的クラスタリングと同様にまず各データをそれぞれクラスタと見なすことに対応し，データ集合に対する圧縮（分割）がまったく行われないことに対応する．この場合には，提案した類似度関数（式 (9)）から導出されるデータグラフにおいて式 (13) で定義される $p(x_j|x_i)$ が定理 1 での条件を満たすことを命題 3 は示している．

4.4　データグラフに基づくアプローチ

本稿では，ハードクラスタリングにおける問題 1 はデータが一様分布する場合には提案するグラフモデルにおける以下の問題に近似できることを示す．なお，問題 1 が問題 2 に近似できるとは，問題 2 に対する解が問題 1 に対する近似解であることを意味する．

問題 2. クラスタ数 k が指定された場合に，データグラフ G において以下の目的関数 J を最小化する互いに素な辺の集合族 $\{\boldsymbol{E}_1, \ldots, \boldsymbol{E}_k\}$ を求めよ．

$$J = \sum_{t=1}^{k} \sum_{w_{ij} \in \boldsymbol{E}_t} w_{ij} \tag{14}$$

ただし，G から $\{\boldsymbol{E}_1, \ldots, \boldsymbol{E}_k\}$ を削除すると G は k 個の連結要素に分割されるものとする．

4.4.1　目 的 関 数

与えられたデータにおいて，データ集合 $\boldsymbol{X}$ とその関連変数 $\boldsymbol{Y}$ に対応する確率変数 X と Y との相互情報量 $I(X;Y)$ はある定数となる．このため，任意の β に対して以下の問題は問題 1 と同値な問題である[19]．

問題 3. 以下の目的関数を最小化する $p(t|x)$ を求めよ．

$$F_{IB} = \sum_x \sum_t p(x)p(t|x)(-\log Z(x, \beta)) \tag{15}$$

ここで，$Z(x, \beta)$ は式 (8) で定義される関数である．

証明（文献 19) 3.1.1 項）

$$\begin{aligned} & \mathcal{L} + \beta I(X;Y) \\ &= I(X;T) + \beta\{I(X;Y) - I(T;Y)\} \\ &= \sum_x \sum_t p(x)p(t|x) \log \frac{p(t|x)}{p(t)} + \beta \sum_x \sum_t p(x)p(t|x) D_{KL}[p(y|x)||p(y|t)] \\ &= \sum_x \sum_t p(x)p(t|x)(-\log Z(x, \beta)) \\ &= F_{IB} \end{aligned} \tag{16}$$

よって，$\mathcal{L}$ の最小化問題（式 (6)）と F_{IB} の最小化問題（式 (15)）は同値な問題である．□

⋆1 各連結要素ごとにデータグラフを考えれば一般性を失わない．

データグラフ G においては，式 (15) の目的関数は以下で表現される．

$$F_G = \sum_{x_i} \sum_{x_j} p(x_i)p(x_j|x_i)(-\log Z(x_i, \beta)) \tag{17}$$

式 (15) と同様，$Z(x_i, \beta)$ は式 (8) で定義される関数である．

頂点 x_i から出る辺の重みの総和を d_i と定義する[⋆1]．

$$d_i = \sum_{x_j} w_{ij}, \quad \forall x_i \in \boldsymbol{X} \tag{18}$$

本稿での主要結果を示すために以下を仮定する．

仮定 1. データは一様分布し，$p(x)$ はある定数 $c > 0$ である．

以下では，この仮定を一様分布と呼ぶ．

命題 4. 一様分布の下では，データ集合$\boldsymbol{X}$に対するデータグラフ G において F_G はある定数である．

証明

$$\begin{aligned} F_G &= \sum_{x_i} \sum_{x_j} p(x_i)p(x_j|x_i)(-\log Z(x_i, \beta)) \\ &= c \sum_{x_i} (-\log Z(x_i, \beta)) \sum_{x_j} p(x_j|x_i) && (19) \\ &= c \sum_{x_i} (-\log d_i) \sum_{x_j} \frac{w_{ij}}{d_i} && (20) \\ &= c \sum_{x_i} (-\log d_i) && (21) \end{aligned}$$

一様分布の下では $p(x_i) = c$（定数）であり，G において各データ x_i ごとに $Z(x_i, \beta)$ はある定数であるため式 (19) が成り立つ．命題 2 と式 (8)，(11)，(18) より $Z(x_i, \beta) = \sum_{x_j} w_{ij} = d_i$ であり，また式 (13) より式 (20) が成り立つ．各 x_i に対して $\sum_j \frac{w_{ij}}{d_i} = 1$ であるため式 (21) が成り立ち，式 (19) と同様に $-\log d_i$ はある定数であるため命題 4 が成り立つ．□

4.4.2　データ圧縮とカット

データ集合 $\boldsymbol{X}$ を$\boldsymbol{X}$= $S \sqcup \bar{S}$ [⋆2]という 2 つの部分集合へ分割することを考える．データグラフ G において部分集合 S と $\bar{S}$ の間を連結する辺を削除することにより，G は対応する誘導部分グラフ G_S と $G_{\bar{S}}$ に分割され，分割後のグラフは（非連結な）グラフ $\hat{G} = \{G_S, G_{\bar{S}}\}$ となる[6]．S と $\bar{S}$ はそれぞれクラスタに対応し，本稿のアプローチは辺の削除によりデータをこれらのクラスタに割り当てるというハードクラスタリングに対応する．

定義 6. 辺の削除による分割に対して $cut(S, \bar{S})$，$cut(\bar{S}, S)$ を以下で定義する．

$$cut(S, \bar{S}) = \sum_{x_i \in S} \sum_{x_j \in \bar{S}} w_{ij} \tag{22}$$

$$cut(\bar{S}, S) = \sum_{x_i \in \bar{S}} \sum_{x_j \in S} w_{ij} \tag{23}$$

命題 5. 各部分集合の要素数が *1* より大きい任意のデータグラフ G の分割に対し，$\frac{w_{ij}}{\sum_{j \in G_S} w_{ij}}$ は各誘導部分グラフ G_S における条件付き確率である．

証明　各 G_S は G の誘導部分グラフであり，$|S| > 1$ のためにそれぞれ連結である．このため，命題 2 より命題 5 が成立する．□

データ集合の分割 $\boldsymbol{X} = S \sqcup \bar{S}$ において，データ集合 $\boldsymbol{X}$ の各要素 x_i に対して x_i を含む部分集合を S_i と表記し，含まない部分集合を $\bar{S}_i$ と表記する．

式 (18) と同様に以下を定義する．

$$d_{S_i} = \sum_{x_j \in S} w_{ij} \tag{24}$$

$$d_{\bar{S}_i} = \sum_{x_j \in \bar{S}} w_{ij} \tag{25}$$

各データ $x_i \in \boldsymbol{X}$ に対し，式 (18)，(24)，(25) の間に以下の関係が成り立つ．

$$d_i = d_{S_i} + d_{\bar{S}_i} \tag{26}$$

命題 5 に基づき，分割後のグラフ $\hat{G}$ における条件付き確率を以下で定義する．

$$\forall x_i \in S, \quad \hat{p}(x_j|x_i) = \begin{cases} \frac{w_{ij}}{d_{S_i}} & x_j \in S \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

$$\forall x_i \in \bar{S}, \quad \hat{p}(x_j|x_i) = \begin{cases} \frac{w_{ij}}{d_{S_i}} & x_j \in \bar{S} \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

式 (17) と同様に F_{G_S} と $F_{G_{\bar{S}}}$ を定義すると，一様分布の下では以下が成り立つ．

⋆1 $\sum_{x_j}$ はすべての $\boldsymbol{X}$ にわたる和であり，$\sum_j$ に対応する．

⋆2 部分集合 $\bar{S}$ は部分集合 S の補集合に対応する．

$$F_{\hat{G}} = \sum_{x_i}\sum_{x_j} p(x_i)\hat{p}(x_j|x_i)(-\log Z(x_i,\beta)) \tag{29}$$

$$= \sum_{x_i\in S}\sum_{x_j\in S} p(x_i)\frac{w_{ij}}{d_{S_i}}(-\log Z(x_i,\beta)) + \sum_{x_i\in \bar{S}}\sum_{x_j\in \bar{S}} p(x_i)\frac{w_{ij}}{d_{S_i}}(-\log Z(x_i,\beta)) \tag{30}$$

$$= F_{G_S} + F_{G_{\bar{S}}} \tag{31}$$

$$= c\sum_{x_i\in S}(-\log d_{S_i}) + c\sum_{x_j\in \bar{S}}(-\log d_{S_j}) \tag{32}$$

ただし，各データ x_i は S_i のみに割り当てられるために $p(S_i|x_i) = 1$，$p(\bar{S}_i|x_i) = 0$，$\forall x_i \in \boldsymbol{X}$ となる．このため，式 (27)，(28) で定義される $\hat{p}(x_j|x_i)$ は式 (7) を満たさず，問題 3 に対する最適解から乖離することになる[⋆1]．

問題 3 を解くために分割にともなう最適解からの乖離を最小化することを考える．命題 4 よりデータ集合 $\boldsymbol{X}$ に対するデータグラフ G において F_G は定数であるため，分割にともなう最適解からの乖離 $F_{\hat{G}} - F_G$ の最小化は以下と同値な問題となる．

問題 4. データグラフ G において，以下を最小化する分割 $\boldsymbol{X} = S \sqcup \bar{S}$ を求めよ．

$$F_{\hat{G}} = F_{G_S} + F_{G_{\bar{S}}} \tag{33}$$

ここで $\hat{G} = \{G_S, G_{\bar{S}}\}$ である．

4.4.3 主 要 結 果

本稿での主要な結果を示す．まず，以下の問題を定義する．

問題 5. データ集合 $\boldsymbol{X}$ に対するデータグラフ G において，以下の目的関数 J_1 を最小化する互いに素な辺の集合族 $\{\boldsymbol{E}_1, \boldsymbol{E}_2\}$ を求めよ．

$$J_1 = \sum_{t=1}^{2}\sum_{w_{ij}\in \boldsymbol{E}_t} w_{ij} \tag{34}$$

ただし，G から $\{\boldsymbol{E}_1, \boldsymbol{E}_2\}$ を削除すると G は 2 個の連結要素に分割されるものとする．

主張 6. ハードクラスタリングにおいては，一様分布の下では問題 *1* は問題 *5* に近似できる．

証明　問題 1 は問題 4 に帰着できるため，問題 4 を問題 5 に近似できることを示す．以下では記号 $\Leftrightarrow$ で同値であることを表記し，記号 $\simeq$ で近似を表記する．

$$\begin{aligned}
&\min F_{\hat{G}} \\
\Leftrightarrow\ &\min\{\sum_{x_i\in S}(-\log d_{S_i}) + \sum_{x_j\in \bar{S}}(-\log d_{S_j})\} \\
\simeq\ &\min\{\sum_{x_i\in S} d_{\bar{S}_i} + \sum_{x_j\in \bar{S}} d_{\bar{S}_j} + \sum_{x_i\in S\sqcup\bar{S}}(1-d_i)\} && (35)\\
\Leftrightarrow\ &\min\{\sum_{x_i\in S} d_{\bar{S}_i} + \sum_{x_j\in \bar{S}} d_{\bar{S}_j}\} && (36)\\
\Leftrightarrow\ &\min\{cut(S,\bar{S}) + cut(\bar{S},S)\} && (37)\\
\Leftrightarrow\ &\min\sum_{t=1}^{2}\sum_{w_{ij}\in \boldsymbol{E}_t} w_{ij} && (38)
\end{aligned}$$

式 (32) より一様分布の下では最初の式が成り立ち，式 (26) の関係に基づいて log 関数を Taylor 展開により第 1 次近似して $(-\log d_{S_i}) \simeq d_{\bar{S}_i} + (1-d_i)$ となるため式 (35) が成り立つ．命題 4 と同様に各 d_i は G において定数であるため式 (36) と同値であり，$cut(S,\bar{S})$ の定義から式 (37) に同値であり，また式 (38) に同値となる．よって，主張 6 が成り立つ．□

主張 6 は以下に拡張できる．

主張 7. ハードクラスタリングにおいては，一様分布の下では問題 *1* は問題 *2* に近似できる．

4.5 データグラフに基づくクラスタリング

前節より，提案するグラフモデルに基づいてデータ集合 $\boldsymbol{X}$ をデータグラフとして表現することにより，3 章での相互情報量に基づくクラスタリング問題をデータグラフにおける組合せ最適化問題（問題 2）の観点からアプローチすることが可能となる．たとえば，この問題を効率的に解く様々なアルゴリズム[13),21)] の利用が考えられる．

ただし，カットに基づくクラスタリングの定式化においては非常に小さなクラスタが生成されてクラスタ集合のサイズに偏りが生じるという問題がある[24)]．データ集合をいくつかのクラスタに分割するというクラスタリングの観点からは，クラスタの偏り（極端にサイズの小さなクラスタの生成など）は望ましいことではないと考えられる．このため，データグラフに基づいて相互情報量に基づくクラスタリング問題を解く際には，目的関数の最小化に加えてクラスタ相互のバランスを考慮することが重要となると考えられる．

⋆1 辺の削除によりハードクラスタリングを行うことに対応する．一般にハードクラスタリングでは最適解から乖離する．

表 1　20 Newsgroup に対するデータセット
Table 1　Datasets from 20 Newsgroup dataset.

データセット	含まれるグループ名
Multi5	comp.graphics, rec.motorcycles, rec.sport.baseball, sci.space, talk.politics.mideast
Multi10	alt.atheism, comp.sys.mac.hardware, misc.forsale, rec.autos, rec.sport.hockey, sci.crypt, sci.med, sci.electronics, sci.space, talk.politics.guns
Multi15	alt.atheism, comp.graphics, comp.sys.mac.hardware, misc.forsale, rec.autos, rec.motorcycles, rec.sport.baseball, rec.sport.hockey, sci.crypt, sci.electronics, sci.med, sci.space, talk.politics.guns, talk.politics.mideast, talk.politics.misc

5. 評　　価

5.1 文書クラスタリングへの適用

先行研究[19] に基づき，提案したグラフモデルを文書クラスタリングに適用して評価した．文書クラスタリングとは文書集合 $\boldsymbol{X} = \{x_1, \ldots, x_n\}$ をクラスタ集合 $\boldsymbol{T}$ に分割する問題であり，各文書 x は文書処理で標準的な単語の頻度に基づくベクトル空間モデルで表現されると仮定する．3.2 節での例と同様，$\boldsymbol{X}$ を表現する全単語集合が $\boldsymbol{Y} = \{y_1, \ldots, y_m\}$ に対応し，$p(x, y)$ は文書 x と単語 y の同時確率に対応する．一般に文書に含まれる単語数は膨大であるため高次元スパース表現なデータをクラスタリングすることに対応する．本稿の手法は分割的クラスタリングに対応するためクラスタ数 $k = |\boldsymbol{T}|$ は与えられると仮定する．

評価対象として，文書クラスタリングのベンチマークである 20 ニュースグループ（以下，20NG）[⋆1]を使用した．先行研究[5),19)] を参考に，本稿では 20NG に対して 5 クラスタ，10 クラスタ，15 クラスタからなる 3 つの母集団を設定し，各母集団に含まれるクラスタからそれぞれ 50 個ずつの文書を非復元抽出してデータセットを作成した．各母集団に含まれるニュースグループを**表 1** に示す．各母集団に対して 10 個ずつ，計 30 個のデータセットを作成した．各データセットに対して porter stemmer [⋆2]を用いて stemming を行い，stop word を除去して相互情報量で上位 2,000 語の単語を選択した．

5.2 実 験 設 定

5.2.1 手　　法

各データセットに対して 4.3 節のデータグラフを作成し，問題 2 に対応する解を求めてクラスタリングを行った．4.5 節で述べたように，データ集合をクラスタリングする際には処理の目的を反映して生成するクラスタ相互のバランスを考慮することが重要になると考えられる．本稿では，この点を考慮した手法としてスペクトルクラスタリングを用いた[24)]．

スペクトルクラスタリングとは，データ間の類似度を表現する非負対称行列 $\mathbf{W}$ が与えられた際，式 (18) で定義される d_i を対角要素とする対角行列 $\mathbf{D}$ から

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \tag{39}$$

を求め，行列 $\mathbf{L}$ に対する l 個の固有ベクトル $\mathbf{H} = \{h_1, \ldots, h_l\}$ を l 次元に埋め込んだ $\boldsymbol{X}$ の表現と見なし，$\mathbf{H}$ で表現されたデータ集合をクラスタリングする手法である．行列 $\mathbf{L}$ はグラフラプラシアンと呼ばれる[1)]．提案するデータグラフでは頂点対 (x_i, x_j) に対してそれぞれ w_{ij}，w_{ji} を持つ辺が定義されるが，分割の際にはその両方を削除する必要がある．このため，以下の実験では式 (11) での重みに基づいて対称行列 $(\mathbf{W})_{ij}$=$(w_{ij} + w_{ji})/2$ を作成した．

クラスタ相互のバランスを考慮するために，対角行列 $\mathbf{D}$ を用いて正規化した以下の 2 つが代表的な手法として提案されている[24)]．

$$\mathbf{L}_{rw} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{W} \tag{40}$$

$$\mathbf{L}_{sym} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}} \tag{41}$$

$\mathbf{L}_{rw}$ は条件付き確率から導出されるグラフ上の酔歩に基づいて正規化したものであり，$\mathbf{L}_{sym}$ は対称に正規化したものである．現状ではどちらを用いるべきかについては一般的な指針はない．このため，それぞれを用いて表現 $\mathbf{H}_{rw}$，$\mathbf{H}_{sym}$ を生成し，これらに対して kmeans 法を用いてクラスタリングを行った．

比較手法として，文献 20)，23) で提案された iIB 法，sIB 法，および高次元スパースデータに対する標準手法である skmeans 法[4)] での実験を行った．iIB 法は式 (7) の定常分布を交互射影により求める手法であり，sIB 法は問題 1 と双対な問題をデータの逐次再割当てにより求める手法である．

なお，テキスト処理でしばしば遭遇するゼロ頻度問題[14)] のために KL 情報量は数値的に不安定となる恐れがある．このため，Ristad 法[17)] でのスムージングを用いて各データセットでの文書 x と単語 y の同時確率 $p(x, y)$ を推定した．

5.2.2 評 価 尺 度

各データセットに対して，各データに対する真のクラスタと各手法が割り当てるクラスタに基づいて以下で述べる正規化相互情報量（NMI）と純度を評価した．

真のクラスタと割り当てられたクラスタに対応する確率変数を T，$\hat{T}$ とすると，正規化

⋆1 http://people.csail.mit.edu/~jrennie/20Newsgroups/．本稿では 20news-18828 を使用した．
⋆2 http://www.tartarus.org/~martin/PorterStemmer

相互情報量（NMI）は以下で定義される．

$$NMI = \frac{I(\hat{T};T)}{(H(\hat{T})+H(T))/2} \quad (\in [0,1]) \tag{42}$$

$H(T)$ はシャノン情報量である．NMI における正規化には様々な手法があるが[22]，本稿では平均による正規化とした．NMI が大きいほど真のクラスタでのデータ割当てに合致することを示す．

純度（purity）は，真のクラスタ C_i と割り当てたクラスタ A_h の分割表に基づいて以下で定義される．

$$purity = \frac{1}{n}\sum_{i=i}^{k}\max_{h}|C_i \cap A_h| \quad (\in [0,1]) \tag{43}$$

n は全データ数であり，クラスタどうしで共有される要素数に基づいて算出される．純度が大きいほど生成されたクラスタのまとまりが高いことを示す．

5.2.3　パラメータ

問題 1 で述べたように，相互情報量に基づくクラスタリングの枠組みにおける主要なパラメータはハイパーパラメータ β であり，式 (9) でも用いられる．sIB 法では $\beta = 10^4$ と非常に大きく設定してハードクラスタリングとすることで β の影響を受けないようにしているが[19]，iIB 法および提案手法では結果は β の値に依存する．このため，予備実験により各手法に対する適切な β の範囲を求め，iIB 法では $\beta \in [1,100]$，提案法では $\beta \in [10^{-2},1]$ として実験した．

スペクトルクラスタリングでは埋め込む次元数 l も影響をおよぼす．基本的に $l = k$（k はクラスタ数）としたが，Multi5 に対しては 5 次元では低次元すぎると考え $l = 10$ とした．

5.3　結　　果

3 つの母集団に対してそれぞれ非復元抽出で 10 個ずつ作成した計 30 のデータセットに対し，初期値依存性を考慮して各データセットごとに 10 回試行を行った．各母集団に対する平均結果を**図 2**，**図 3** に示す．図 2，図 3 ではグラフモデルに $\mathbf{L}_{rw}$ を用いたものを kl-rw，$\mathbf{L}_{sym}$ を用いたものを kl-sym と表記した．上記のように β が主要なパラメータであるため，各図ごとに横軸に β，縦軸に評価値とした．ただし，sIB 法では文献 19) に従って各データセットに対する 10 回試行での最良値から平均を計算した．

各データのクラスタへの割当ての正しさに対応する NMI に関しては（図 2），Multi10，Multi15 に対して提案したグラフモデルに $\mathbf{L}_{rw}$ を用いた手法（kl-rw）が他手法より大きな値であった．他方，Multi5 に対しては，iIB 法や skmeans 法より大きな値であったが，sIB 法よりは小さかった．

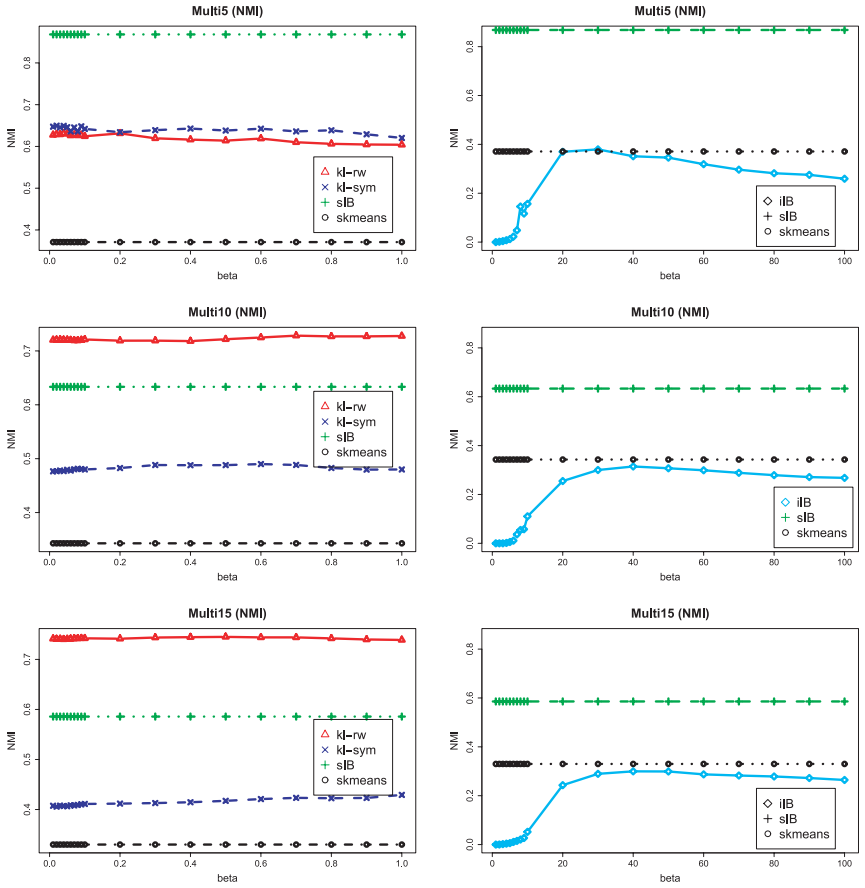

図 2　20NG に対する結果（NMI）
Fig. 2　Result on 20NG (w.r.t. NMI).

図 3 に示すように，クラスタのまとまりに対応する純度に関しても NMI に関する結果とほぼ同様の結果となった．Multi10，Multi15 に対して提案したグラフモデルに $\mathbf{L}_{rw}$ を用いた手法（kl-rw）が他手法を上回ったが，Multi5 に対しては iIB 法や skmeans 法を上回った

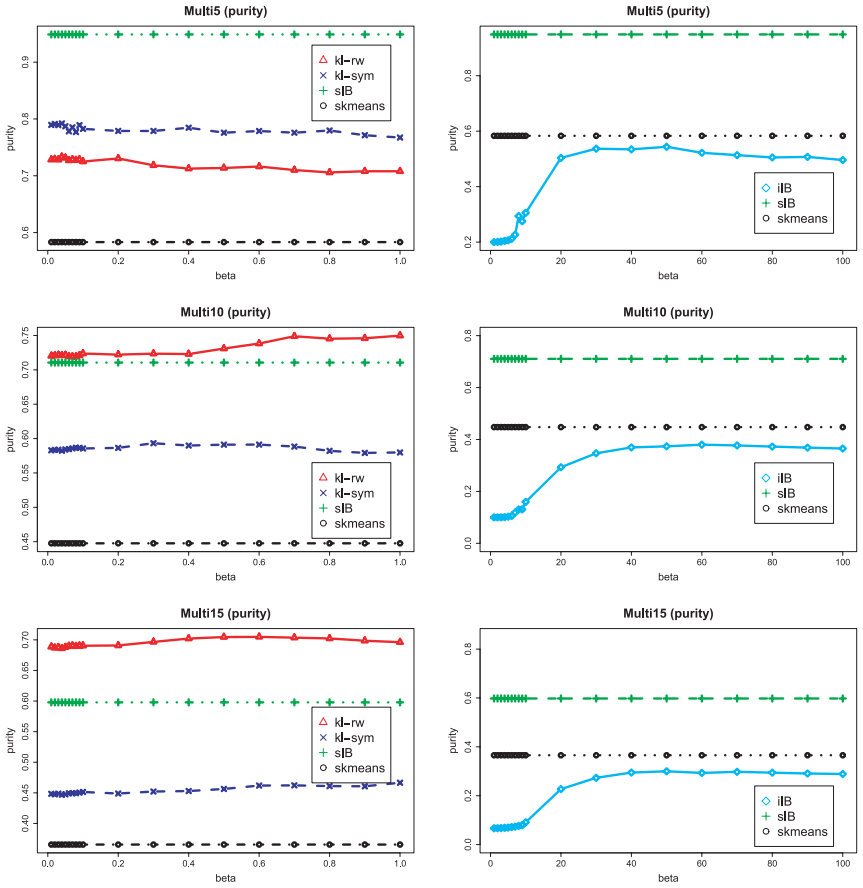

図 3　20NG に対する結果（purity）
Fig. 3　Result on 20NG (w.r.t. purity).

ものの sIB 法にはおよばなかった．

6. 議　　論

6.1 グラフモデルの妥当性

本稿での提案は，3 章で述べた相互情報量に基づくクラスタリングの枠組みに対し，定理 1 での定常分布に着想を得たデータ間の類似度関数（式 (9)）を定義してデータ集合をデータグラフとして表現することにより，辺重み付きグラフ上の問題に帰着させて問題 1 を解くというアプローチである．定理 1 での定常分布を求めるという意味では iIB 法に対応すると考えられるが，5.3 節での結果より iIB 法との比較を通じて提案手法の妥当性と有効性を確認した．

5.2.1 項で述べたように $\mathbf{L}_{rw}$ と $\mathbf{L}_{sym}$ のどちらを用いるべきかについては一般的な指針はないが，本稿のアプローチはデータグラフ上の重みから導出される条件付き確率に基づいて問題 1 をグラフ上の問題に帰着させたものである．このため，条件付き確率から導出されるグラフ上の酔歩に基づく $\mathbf{L}_{rw}$ のほうが提案するグラフモデルに合致すると考えられる．図 2，図 3 の結果からもクラスタ数が増加した場合には kl-rw が kl-sym を NMI と純度の両方で上回り $\mathbf{L}_{rw}$ が提案モデルに合致することが確認できるため，グラフモデルは相互情報量に基づくクラスタリング問題のモデルとして妥当であると考えられる．

本稿のアプローチは，式 (6) で定義される情報論的な目的関数をグラフモデルに基づいてグラフカットで近似するというアプローチであるが，グラフカットによる近似は他分野でも用いられる．たとえば画像復元をエネルギー最小化に基づいて行う際にもグラフカットを用いて高速に近似解を求める手法が提案されている[11]．4.4.1 項で述べたように問題 1 と問題 3 は同値な問題であり，式 (15) は統計物理における自由エネルギーに相当するため，本稿のアプローチもエネルギー最小化を実現する際にグラフカットを用いることで高速に近似解を求めるものと見なすことができる．

式 (35) において log 関数を近似する際にはテーラー展開による一時近似 $\log(1+x) = x$ を用いたが，式 (24) で定義される d_{S_i} の大きさは与えられたデータに依存するため，この近似に対する一般的な上界を示すことは現状ではできていない．グラフモデルにおいてグラフカット問題に変換する際の近似に対するより詳細な議論は今後の課題である．

また，式 (40)，(41) での $\mathbf{L}_{rw}$ や $\mathbf{L}_{sym}$ による正規化はクラスタ相互のバランスを考慮するためのものであり，式 (34) の目的関数に対してクラスタリングの観点から用いたものである．今後は，式 (34) を拡張して正規化も含めた目的関数をモデル化することに取り組む予定である．

6.2 グラフモデルに基づくクラスタリング

クラスタの質の評価に関して様々な指標が提案されているが[8]，クラスタ割当ての正しさに対応する NMI とまとまり具合に対応する純度の観点からの評価においてグラフモデルに $\mathbf{L}_{rw}$ を用いた場合はクラスタ数が多い場合に sIB 法を上回る結果を示した．このため，提

表 2　スペクトルクラスタリングとの比較（NMI）
Table 2　Comparison with Spectral Clustering (NMI).

データセット	$\mathbf{L}_{rw}$	$\mathbf{L}_{sym}$	提案+$\mathbf{L}_{rw}$ ($\beta = 10^{-2}$)
Multi5	0.573	0.641	0.627
Multi10	0.534	0.497	0.720
Multi15	0.464	0.424	0.741

案法は有効なクラスタリング手法であると考えられる．

グラフモデルに基づいてクラスタリングを行う際にはスペクトルクラスタリングを用いた．スペクトルクラスタリング自体の影響を調べるため，5.2.2 項のデータを文書クラスタリングで一般的に用いられるコサイン類似度を用いて各データセットをグラフとして表現し，同じ実験設定のもとで式 (40)，(41) の正規化に基づくスペクトルクラスタリングを適用した結果を**表 2** に示す．表 2 より提案法は上記のスペクトルクラスタリング法を大きく上回る性能を示しており，この結果は提案モデルの有効性を示していると考えられる．

提案法はスペクトルクラスタリングと同様にデータどうしの関係に基づいてデータ集合 $\boldsymbol{X}$ をグラフとして表現するため，まずグラフを構築する必要がある．$|\boldsymbol{X}| = n$ とするとグラフ構築には $O(n^2)$ を要し，固有ベクトルの計算（l 本の固有ベクトルの計算は条件にもよるが $O(ln)$ で可能[25]）よりも計算コストが高い[⋆1]．他方，逐次再割当てに基づく sIB や skmeans は収束までのループ回数にも依存するが $O(kn)$（k はクラスタ数）であり，計算コストの面では有利である．

6.3　文書クラスタリングへの応用

高次元スパースデータという困難な問題である文書クラスタリングに対して，提案手法（グラフモデルに $\mathbf{L}_{rw}$ を用いた場合）は NMI および純度に関して Multi10，Multi15 において sIB 法をも上回る性能を示し，クラスタ数が多い場合にクラスタへの割当ての正しさという観点からの有効性を確認した．しかし，Multi5 に対しては残念ながら sIB 法にはおよばなかった．この理由として，問題 1 は KL 情報量に基づく相互情報量により定式化されているが，文書クラスタリングへの適用に際してはテキスト処理でしばしば遭遇するゼロ頻度問題[14] のために数値的に不安定となることが考えられる．表現に用いる単語数を絞り込むことによりこの問題を回避することも試みたが，現状では Multi5 のようにクラスタ数が少ない場合には残念ながら sIB 法にはおよばなかった．実データへの適用に際して上記の問題に対処することは今後の課題である．

提案モデルは適用対象を特に限定するものではないが，先行研究[19] に基づいて文書クラスタリングへの適用結果を報告した．しかし，現状のモデルにおける一様分布の妥当性は適用分野やデータに依存し，また文書クラスタリングにおいては文書クラスの大きさの分布を考慮することが重要になるとの知見もある．今後は，文書処理の特性を反映したモデル化や目的関数の設定に取り組む予定である．

⋆1 ただし，グラフ構築や固有ベクトルの計算は並列処理による高速化も可能である．

7. おわりに

本稿では，相互情報量に基づくクラスタリング問題に対するグラフモデルを提案し，データが一様分布する場合にはハードクラスタリング問題が提案するグラフモデルにおける組合せ最適化問題に近似できることを示した．相互情報量から導出される定常分布に着想を得たデータ間の類似度関数を定義してデータ集合全体を辺重み付きグラフとして表現し，もとの問題とグラフ上の組合せ問題との対応を示した．

提案するグラフモデルを用いてデータ集合を表現することにより，グラフ構造に基づく様々なアルゴリズムを用いて相互情報量に基づくクラスタリングを行うことが可能になると考えられる．一例として，文書クラスタリングでのベンチマークデータである 20 Newsgroup のデータをグラフモデルで表現し，スペクトルクラスタリング法を適用して評価し，他手法との比較を通じて提案手法の妥当性と有効性を確認した．しかし，文書クラスタリングなどのように高次元スパースな実データに対して適用する際にはテキスト処理におけるゼロ頻度問題にともなう問題に対処することが重要となることも判明した．今後はこの問題に取り組む予定である．

謝辞　本研究の一部は文部科学省科研費（No. 20500123）の補助による．最後に，有益なご指摘を賜りました査読者の方々に深く謝意を表します．

参 考 文 献

1) Chung, F.: *Spectral Graph Theory*, American Mathematical Society (1997).
2) Cover, T. and Thomas, J.: *Elements of Information Theory*, Wiley (2006).
3) Dempster, A., Laird, N. and Rubin, D.: Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society, Series B* (*Methodological*), Vol.39, No.2, pp.1–38 (1977).
4) Dhillon, J. and Modha, D.: Concept Decompositions for Large Sparse Text Data

using Clustering, *Machine Learning*, Vol.42, pp.143–175 (2001).
5) Dhillon, J., Mallela, S. and Modha, D.: Information-theoretic co-clustering, *KDD 2003*, pp.89–98 (2003).
6) Diestel, R.: *Graph Theory*, Springer (2006).
7) Elghazel, H., Yoshida, T., Deslandres, V., Hacid, M. and Dussauchoy, A.: A New Greedy Algorithm for improving b-Coloring Clustering, *Proc. 6th Workshop on Graph-based Representations*, pp.228–239 (2007).
8) Ghosh, J.: *Scalable clustering*, pp.341–364, Lawrence Erlbaum Assoc. (2003).
9) Guënoche, A., Hansen, P. and Jaumard, B.: Efficient algorithms for divisive hierarchical clustering with the diameter criterion, *Journal of Classification*, Vol.8, pp.5–30 (1991).
10) Hartigan, J. and Wong, M.: Algorithm AS136: A k-means clustering algorithm, *Journal of Applied Statistics*, Vol.28, pp.100–108 (1979).
11) 石川　博：チュートリアル「グラフカット」，情報処理学会研究報告，Vol.2007-CVIM-158, No.26, pp.193–204 (2007).
12) Jain, A., Murty, M. and P.J., F.: Data Clustering: A Review, *ACM Computing Surveys*, Vol.31, pp.264–323 (1999).
13) Kamidoi, Y., Yoshida, N. and Nagamochi, H.: A Deterministic Algorithm for Finding All Minimum k-Way Cuts, *SIAM Journal on Computing*, Vol.36, No.5, pp.1329–1341 (2006).
14) 北　研二：確率的言語モデル，東京大学出版会 (1999).
15) Ng, R. and Han, J.: CLARANS: A method for clustering objects for spatial data mining, *IEEE Trans. Knowledge and Data Engineering*, Vol.14, No.5, pp.1003–1016 (2002).
16) Pereira, F., Tishby, N. and Lee, L.: Distributional clustering of English words, *Proc. 30th Annual Meeting of the Association for Computational Linguistics*, pp.183–190 (1993).
17) Ristad, E.: A Natural Law of Succession, Technical Report CS-TR-495-95, Princeton University (1995).
18) Slonim, N. and Tishby, N.: Agglomerative Information Bottleneck, *Advances in Neural Information Processing Systems* (*NIPS*) *12*, pp.617–623 (1999).
19) Slonim, N.: The Information Bottleneck: Theory and Applications, Ph.D. Thesis, Hebrew University (2002).
20) Slonim, N., Friedman, N. and Tishby, N.: Unsupervised Document Classification using Sequential Information Maximization, *SIGIR-02* (2002).
21) Stoer, M. and Wagner, F.: A Simple Min-Cut Algorithm, *J. ACM*, Vol.44, No.4, pp.585–591 (1997).
22) Strehl, A. and Ghosh, J.: Cluster Ensembles —A Knowledge Reuse Framework for Combining Multiple Partitions, *J. Machine Learning Research*, Vol.3, No.3, pp.583–617 (2002).
23) Tishby, N., Pereira, F. and Bialek, W.: The Information Bottleneck Method, *Proc. 37th Allerton Conference on Communication and Computation* (1999).
24) von Luxburg, U.: A Tutorial on Spectral Clustering, *Statistics and Computing*, Vol.17, No.4, pp.395–416 (2007).
25) 山本有作：密行列固有値解法の最近の発展（I）：Multiple Relatively Robust Representations アルゴリズム，日本応用数理学会論文誌，Vol.15, No.2, pp.181–208 (2005).

(平成 21 年 11 月 18 日受付)
(平成 21 年 12 月 31 日再受付)
(平成 22 年 1 月 26 日採録)

吉田　哲也（正会員）

1968 年生．1991 年東京大学工学部航空工学科卒業．1992 年から 1993 年にかけてエジンバラ大学大学院留学．1997 年東京大学大学院博士課程修了．工学博士．同年大阪大学大学院基礎工学研究科助手．2001 年大阪大学産業科学研究所助手．2004 年北海道大学大学院情報科学研究科助教授．現在，同大学准教授．主に機械学習，知識獲得，データマイニング等の研究に興味を持つ．人工知能学会会員．