

教師ありクラスタリングと絶対/相対クラスタリング Supervised Clustering and Absolute/Relative Clustering

神嶋 敏弘*

Toshihiro Kamishima

Abstract: Researches concerning (semi-)supervised clustering are recently emerging. We show two types of clustering tasks which should be axiomatically differentiated under this supervised setting.

keywords: (Semi-)Supervised-Clustering, Reference Matching, Noun Phrase Coreference

1 はじめに

クラスタリングとは、対象集合をクラスタに分割することである。しかし、このクラスタは内的結合や外的分離といった曖昧な概念によって定義されているため、クラスタリングは well-defined な問題ではない。このことは、しばしば指摘されてきた [12, 13]。そこで、望ましいクラスタが満たすべき条件の例題を教師情報として与え、その情報に基づいて分割を行う**教師ありクラスタリング**が、近年、研究されている。この教師ありクラスタリングの枠組みを用いた場合、獲得すべき「真の分割」には公理的に区別すべき絶対/相対クラスタリングの2種類があると文献 [18] で述べた。ここでは、この絶対/相対クラスタリングと教師ありクラスタリングとの関連について議論を深める。

2 と 3 節では絶対/相対クラスタリング問題と具体例について、4 節では教師ありクラスタリングの分類、そして 5 節では、教師ありクラスタリングと絶対/相対クラスタリングとの関連について述べ、6 節でまとめを述べる。

2 絶対/相対クラスタリング

文献 [18] の絶対/相対クラスタリングの定義と関連事項を示す。全ての可能な対象の集合 \mathcal{X} の任意の部分集合を $X = \{\mathbf{x}_i\}_i^N$ で表す。 X の分割とは、 X について網羅的で、互いに素な X の部分集合 c_k の集合 $C = \{c_k\}_k^K$ であり、これら部分集合 c_k をクラスタと呼ぶ。クラスタリング関数 $\pi(X)$ はクラスタリング問題を表現し、与えられた X に対して、 X の可能な分割の全体の集合から適切な分割を出力する。関数 $\delta(\{\mathbf{x}_i, \mathbf{x}_j\}, C)$ は、 $i \neq j$

なる対象 \mathbf{x}_i と \mathbf{x}_j が、 C 中の同一クラスタの要素なら 1 をとり、そうでなければ 0 をとる関数である。

定義 1 クラスタリング関数 $\pi(X)$ が式 (1) の条件を満たすなら、この関数が表すクラスタリング問題を**絶対クラスタリング (absolute clustering)**、そうでないなら**相対クラスタリング (relative clustering)**という。

$$\delta(\{\mathbf{x}_i, \mathbf{x}_j\}, \pi(X)) = \delta(\{\mathbf{x}_i, \mathbf{x}_j\}, \pi(X')), \\ \forall \mathbf{x}_i, \mathbf{x}_j \in X \cap X', \mathbf{x}_i \neq \mathbf{x}_j, \forall X, X' \subseteq \mathcal{X} \quad (1)$$

すなわち、対象 \mathbf{x}_i と \mathbf{x}_j が同じクラスタに分類されるかどうかは、クラスタリングする対象集合とは独立で、 \mathbf{x}_i と \mathbf{x}_j のみに依存して決まるなら絶対クラスタリングである。絶対クラスタリングであるの必要十分条件は、次の絶対分割が存在することである。

定義 2 式 (2) の条件を満たす \mathcal{X} の分割を絶対分割 $C = \pi(\mathcal{X})$ という。

$$\delta(\{\mathbf{x}_i, \mathbf{x}_j\}, \pi(X)) = \delta(\{\mathbf{x}_i, \mathbf{x}_j\}, C), \\ \forall \mathbf{x}_i, \mathbf{x}_j \in X, \mathbf{x}_i \neq \mathbf{x}_j, \forall X \subseteq \mathcal{X} \quad (2)$$

式 (1) が成立するなら $X' = \mathcal{X}$ とすれば式 (2) になる。逆に、任意の X と X' について式 (2) が成立するので、 $\delta(\{\mathbf{x}_i, \mathbf{x}_j\}, \pi(X)) = \delta(\{\mathbf{x}_i, \mathbf{x}_j\}, C) = \delta(\{\mathbf{x}_i, \mathbf{x}_j\}, \pi(X'))$ となり式 (1) も成立する。よって、絶対分割が存在すれば絶対クラスタリングで、存在しなければ相対クラスタリングである。

3 絶対/相対クラスタリングの例

教師ありクラスタリングが適用されてきた具体例を示し、それらが絶対/相対クラスタリングのどちらであるかを述べる。

*産業技術総合研究所, National Institute of Advanced Industrial Science and Technology (AIST), <http://www.kamishima.net/>

3.1 Reference Matching 問題

同一の文献や人物を同じクラスタに分類する reference matching 問題について述べる [10, 30]. 他の論文を引用する場合, 同じ文献の引用であっても, 違った書式で書かれるため, 同一の文献かどうかの判定は困難な場合がある. 例えば, 「International Conference on Machine Learning」は「ICML」とも書かれる. 「T.Kamishima」と「神嶋敏弘」も同一の人物である. さらに「著者→題名→…」の順や「著者→年→…」の順で書かれたりする. reference matching 問題とは, このように違った書式の文献や人物から, 同一のものを特定する問題である. より一般的に, データベース中のレコードで, 現実世界の同じ実体を示すものをまとめる問題は record linkage や identity uncertainty と呼ばれる.

このクラスタリング問題では, 任意の参照文字列 A と B が, 真に同じ文献を示すものであれば, それらは同時に分類する他の文字列とは無関係に同じクラスタに分類すべきである. よって, この問題は絶対クラスタリングである. 全ての可能な文献の参照文字列の集合は \mathcal{X} に相当し, 絶対分割の各クラスタは同一の文献を参照する全ての文字列で構成されている.

3.2 名詞句の coreference 問題

名詞句の coreference とは, 現実世界で同じ対象を表す文書中の名詞句を同じクラスタに分類する問題である [10, 16, 28]. 例えば, 総理大臣の動向の記事を考える. その中で「安倍氏」「首相」「総理」「彼」などの表現は全て同一の人物を指し示すと見なせるので, これらの語を同じクラスタに分類する. この問題も, 語の間に類似度を適切に定め, クラスタリング手法を適用して解かれてきた.

この問題は相対クラスタリングであるが, その理由を述べる. 次の文 A~C から構成される文書を考える.

- A. 親亀がいる.
- B. この亀に子亀が乗っている.
- C. この亀に孫亀が乗っている.

文 B と文 C の「この亀」を, それぞれ (B) と (C) により区別する. coreference 問題を考えて, 生成されるべきクラスタは {親亀, この亀 (B)}, {子亀, この亀 (C)}, そして {孫亀} である. 「親亀」と「この亀 (C)」は, このとき, 異なるクラスタの要素である. ここで, 文 B を取り去ると「この亀 (B)」と「子亀」の二つの名詞句が対象集合に含まれなくなる. すると, 生成されるべきクラスタは {親亀, この亀 (C)} と {孫亀} に変化し, 含まれなくなった二つの名詞句の影響を受けて「親亀」と「この亀 (C)」

は同じクラスタに分類されるようになる. すなわち, 二つの名詞句が同じクラスタに分類されるかどうかは, これら二つの名詞句だけでは決定できない. よって, 名詞句の coreference 問題は相対クラスタリングである.

3.3 その他の問題

画像の領域分割とは, 画素などの基本構成要素を, 同じ対象物を表すものが一つのクラスタになるようにする問題である [19]. この問題では, 二つの画素の間や周囲にある他の画素の影響されて, 二つの画素が同じ領域になるかどうかが決まる. よって画像の領域分割は相対クラスタリング問題である.

同じカテゴリの文書がクラスタを構成するように分類する文書クラスタリングはやや複雑である. もし, 未知だが不変な意味的なカテゴリ分類があって, そのカテゴリに従ってクラスタを生成するなら絶対クラスタリング問題となる [5, 22, 25, 29]. 一方, 与えられた文書集合を簡潔に二つに分割したいとする. このとき, 文書集合が今週と先週の新聞記事で構成されているなら, 掲載された週で分類し, 政治と経済の記事で構成されているなら, これらのカテゴリによって分類するのが目標だとする. この場合は, 一对の記事が同じクラスタになるかは, 文書集合の構成要素に依存するので, 相対クラスタリング問題となる.

文中の述語を, 同じ意味の用例ごとにまとめる問題 [34] では, 新たな述語を加えても同じ用例に分類された述語は同じクラスタに分類されたままであるべきなので, 絶対クラスタリング問題である.

DNA から転写された mRNA 前駆体を, 符号領域である exon とそうでない intron に切断する splicing の予測問題もクラスタリング問題と見なせる. 一对の塩基が同じ exon や intron の領域になるかどうかは, 周囲の塩基に依存するので, この問題は相対クラスタリングである.

4 教師ありクラスタリング

教師ありクラスタリングの定義は, 個々の研究者がそれぞれに与えているため, 現状では非常に混乱している. よって, 本節ではこの論文中での便宜上の分類を示す.

最初に, クラス分類問題, 特に, ラベルあり/なし混在データから学習する場合との区別を述べる. これはラベルが付加されたデータに加え, ラベルがないデータも同時に用いてクラス分類器を獲得する問題である. この問題を準教師ありクラスタリングと呼ぶこともあるが, Chang ら [7] が提案するように準教師ありクラス分類 (semi-supervised classification) と呼ぶことにする. クラスラベルの数が有限で, 全てのクラスラベルが

既知であるならクラス分類問題、そうでなければクラスターリングとする。具体的には、ラベルの形式で教師情報を与えていても、それ以外の未知のラベルが存在する場合はクラスターリングとみなす。逆に、文献 [4] のように、ラベルあり/なし混在データをクラスターリング手法で処理していても、ラベルが全て既知であれば、準教師ありクラス分類とする。

次に、**制約付クラスターリング (constrained clustering)** と教師ありクラスターリングを区別する。与えられた制約・教師情報の一般化がなされない、すなわち情報が示されている対象だけについて適用されるなら制約付きクラスターリング、でなければ教師ありクラスターリングとする。Wagstaff らの COP-KMEANS[35] を例に説明する。この方法では、must リンクと cannot リンクの二種類の制約条件がある。must リンクで結ばれた対象の対は同じクラスタの要素に、cannot リンクで結ばれた対象の対は違うクラスタの要素になければならない¹。これらの制約の下で、 k -means 法でクラスターリングをする。この方法では、制約の一般化は行われず、制約された対象のどんなに近くにある対象でも、制約を受けることはない。よって、ここでは制約付きクラスターリングに分類する。制約には、同じクラスタの要素が隣接関係の連結成分でなければならない制約 [15] や、障害物といった制約 [14, 33] などもある。あと、初期化などに副次的に教師情報を利用する手法 [1, 37] もあるが、これらは教師なしクラスターリングとして扱う。

教師ありクラスターリング (supervised clustering) を、ラベル集合が事前には未知であるような分類問題であるクラスターリングを、与えられた教師情報を一般化しつつ利用して行うことと定義する。この教師ありクラスターリングを、さらに、**完全教師ありクラスターリング (fully supervised clustering)**、**準教師ありクラスターリング (semi-supervised clustering)**、および **transductive クラスターリング** に分類する。完全教師ありクラスターリングでは、対象集合 $X_i \subset \mathcal{X}$ と、 X_i に関して正当なクラスタについての教師情報 Y_i (次節で詳細に述べる) の対の集合 $\{(X_i, Y_i)\}_i^N$ を訓練事例とする。一方、準教師ありクラスターリングでは、訓練事例は一つの対象集合 $X \subset \mathcal{X}$ と、この集合に関して正当な教師情報 Y との対を訓練事例とする。完全教師ありも準教師ありもどちらの場合でも、テスト対象集合 $X_t \subseteq \mathcal{X}$ が与えられたとき、適切な分割 \hat{C}_t を出力するクラスターリング関数 $\pi(X_t)$ を、訓練事例から獲得することが目標である。最後の transductive クラスターリングは、準教師

ありクラスターリングと同じ形式の訓練事例 (X, Y) を扱う。だが、 X の適切な分割 C を求めることが目的で、未知の対象を分類するクラスターリング関数を求めることはしない。よって、新たな対象集合 X' については、準教師ありの場合と異なり、再学習が必要になる。これは、リッジ回帰を対象にした transductive 学習 [8] と共通性があるので transductive クラスターリングと呼ぶ。

整理すると、まず、完全なクラスの定義が事前に与えられているかによってクラス分類とクラスターリングを区別する。クラスターリングでは、制約や教師情報を一般化するかどうかで、制約付クラスターリングと教師ありクラスターリングとを分ける。教師ありクラスターリングを、訓練事例の形式と学習の目標によって、完全教師あり、準教師あり、および transductive クラスターリングに分ける。最後に、文献 [7] の提案を採用し、教師ありクラスターリングと準教師ありクラス分類を併せて、準教師あり学習と呼ぶ。

4.1 教師ありクラスターリングの教師情報

教師ありクラスターリングの教師情報 Y について述べる。完全教師ありクラスターリングでは、教師情報 Y_i は X_i の正しい分割 C_i が用いられてきた [10, 16, 19, 20, 21]。準教師ありの場合には、 X の正しい分割 C も、教師情報 Y として利用できる [26, 29]。だが、transductive の場合には、正しい分割を求めることが目的なので、この形式の教師情報はありえない。

準教師ありや transductive の場合では、上記の must リンクや cannot リンク制約を使う手法が多い [5, 6, 7, 9, 17, 22, 23, 24, 27, 28, 36, 38]。同じクラスタになるべき対象の部分集合を与える方法 [2, 3, 11] もあるが、これらの集合内の要素間に must リンク制約を与えるのと等価である。逆に cannot リンクのみを考えるものもある [30]。また、ラベル情報を部分的に与えることも考えられるが、同じ (違う) ラベルが付加された事例間に must(cannot) リンクを想定するのと等価である。さらに、3点 x_i, x_j, x_k について、 x_i と x_j は、 x_i と x_k より同じクラスタになりやすい相対比較 (relative comparison) 制約 [25, 31] もある。他に、分割する対象と関連がある副情報を教師情報とする Tishby らの情報ボトルネック [32] など教師ありクラスターリングといえるだろう。

4.2 教師ありクラスターリングの手法の分類

ここでは、教師ありクラスターリング手法をラッパー型と統合型に分ける。ラッパー (wrapper) 型では、教師情報にあった分割が得られるように、元の入力を変換したものを、既存の教師なしクラスターリングへ入力して分割

¹ must リンクの扱いは容易だが、cannot リンクは制約を満たす分割が可能かを判断するだけでも NP 完全である [23]。

を得る。教師なしクラスタリングへの入力形式は、各対象を記述した特徴ベクトルか、二対象間の類似度を要素とする類似度行列である。だが、提案されている教師ありクラスタリングはどれも類似度を求める関数を学習する [2, 3, 7, 9, 11, 23, 26, 29, 30, 31, 36]。これらの方法は準教師ありクラスタリングを対象としている。クラスタリング関数は、学習した類似度関数と教師なしクラスタリングによって構成される。

もう一つの統合 (unified) 型では、制約を充足し、かつ、クラスタリング本来の内的結合と外的分離を満たすと高評価になる目的関数を利用する。transductive の場合には、この目的関数を最適化する分割を求め、完全教師ありや準教師ありの場合には、目的関数を最適化するクラスタリング関数を求める。この統合型の手法には、完全教師ありの [10, 16, 19, 20, 21] や、transductive の [5, 6, 17, 22, 24, 25, 27, 28, 34, 38] が提案されている。

4.3 教師ありクラスタリング手法

教師ありクラスタリング手法を幾つか挙げておく。

Xing らの方法はラッパー型の準教師ありクラスタリングである。 \mathbf{x}_i と \mathbf{x}_j の類似度 $\|\mathbf{x}_i - \mathbf{x}_j\|_A$ を、半正定値行列 A を用いて $\sqrt{(\mathbf{x}_i - \mathbf{x}_j)^\top A (\mathbf{x}_i - \mathbf{x}_j)}$ で表す。また、教師情報には must リンクと cannot リンクを用い、これらのリンクの集合を S と D と記す。学習は次式を最大化する行列 A を求めることで行う。

$$\begin{aligned} \min_A \sum_{\{\mathbf{x}_i, \mathbf{x}_j\} \in S} \|\mathbf{x}_i - \mathbf{x}_j\|_A^2 & \quad (3) \\ \text{s.t.} \sum_{\{\mathbf{x}_i, \mathbf{x}_j\} \in D} \|\mathbf{x}_i - \mathbf{x}_j\|_A \geq 1, & \quad A \text{ は半正定値} \end{aligned}$$

A の学習後は、テスト対象集合 X_t 中の対象間の類似度が計算でき、その類似度に教師なしクラスタリングを適用して推定分割 \hat{C} を求める。

神鷹ら [19, 20] の方法は、統合型の完全教師ありクラスタリングである。教師情報 Y_i には、 X_i に対する適切な分割 C_i を用いる。対象集合は以下の四種類の属性ベクトルで記述される。

- 分類対象属性 \mathbf{x}_i : X_i 中の各対象の特徴 (点の位置) を表現する。
- 分類対象対属性 \mathbf{p}_{ij} : 対象 \mathbf{x}_i と \mathbf{x}_j の対の特徴 (点間の距離など) を表す。
- クラスタ属性 \mathbf{c}_k : 分割後の各クラスタの特徴 (クラスタ内の要素数など) を表す。
- 分割属性 \mathbf{C} : 分割後の分類対象集合全体の特徴 (クラスタ数など) を表す。

学習段階では、これらの属性の結合確率分布を学習する。

$$\Pr[C=C^*, \{\mathbf{x}_i\}, \{\mathbf{p}_{ij}\}, \{\mathbf{c}_k\}, \mathbf{C}] \quad (4)$$

ただし、 $C=C^*$ は、ある分割 C が真の分割 C^* と一致するという事象である。この関数を $\Pr[C=C^*|\{\mathbf{x}_i\}, \{\mathbf{p}_{ij}\}]$, $\Pr[\{\mathbf{c}_k\}|C=C^*]$, および $\Pr[\mathbf{C}|C=C^*]$ の積に分解して求める。これらを MAP 推定で求めるので、この方法を SC-MAP と呼ぶ。学習した式 (4) を最大化する、 X_t の分割を欲張り法で探索して推定分割 \hat{C}_t とする。

Kulis らの SS-kernel-kmeans[24] は、統合型の transductive クラスタリングである。教師信号には must リンクと cannot リンクを利用し、次式を最小化するように X を分割する。

$$\begin{aligned} \sum_{k=1}^K \sum_{\mathbf{x}_i, \mathbf{x}_j \in c_k} \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{|c_k|} \\ - \sum_{k=1}^K \sum_{\substack{\mathbf{x}_i, \mathbf{x}_j \in c_k \\ \{\mathbf{x}_i, \mathbf{x}_j\} \in S}} \frac{2w_{ij}}{|c_k|} + \sum_{k=1}^K \sum_{\substack{\mathbf{x}_i, \mathbf{x}_j \in c_k \\ \{\mathbf{x}_i, \mathbf{x}_j\} \in D}} \frac{2w_{ij}}{|c_k|} \end{aligned} \quad (5)$$

ただし、 S と D は、それぞれ must リンクと cannot リンクの集合。 w_{ij} は罰則重みで、must リンクが満たされたときには目的関数から減算され、cannot リンクに違反すれば加算される。 $\|\mathbf{x}_i - \mathbf{x}_j\|$ はユークリッドノルムであり、カーネルを用いて計算できる。

5 教師ありクラスタリングと絶対/相対クラスタリング

教師ありクラスタリングと絶対/相対クラスタリングとの関連について議論する。

まず、transductive な場合について考察する。絶対/相対クラスタリングの違いは、クラスタリングする対象集合の構成が変化する場合に生じる。例えば、3.2 節の名詞句の coreference 問題では、文を対象集合から取り除くと、単語のクラスタへの割り当てが変化するため、相対クラスタリングとなった。しかし、transductive クラスタリングの目標は、与えられた対象集合 X を分割することで、要素の追加や削除は考慮しない。よって、クラスタリング問題が絶対か相対であるかは、transductive クラスタリングの場合には考慮する必要がない。

以下、完全教師ありと準教師ありクラスタリングの場合に、絶対/相対クラスタリングの違いが、(1) 訓練サンプルの提示方法と、(2) 対象集合の記述方法の二点に与える影響について議論する。

5.1 訓練サンプルの提示方法

完全教師ありと準教師ありクラスタリングでは訓練サンプルの形式が異なる。完全教師ありの場合は、対象集合 X_i と教師情報 Y_i の対の集合 $\{(X_i, Y_i)\}$ だが、準教師ありの場合には、一個の対 (X, Y) である。ここで、絶対/相対クラスタリングを区別する重要な点を繰り返すと、一対の対象が同じクラスタになるかどうか、対象集合の他の要素に依存するかどうかということである。すなわち、相対クラスタリング問題では、各教師情報 Y_i は、対応する対象集合 X_i に依存している。そのため、任意のテスト対象集合 X_i を分割できるクラスタリング関数を獲得するには、様々な対象についての教師情報が与えられている完全教師ありのアプローチを採用すべきである。

逆に、絶対クラスタリング問題では必ず準教師ありクラスタリングのアプローチを、次の理由により、採用すべきである。絶対クラスタリング問題で、完全教師あり型のサンプル集合があったとする。ここで、対象集合 X_1 と X_2 があり、 X_1 中の $\mathbf{x}_i, \mathbf{x}_j$ 間と X_2 中の $\mathbf{x}_j, \mathbf{x}_k$ 間に must リンク制約があったとしよう。式 (1) の条件から、対象集合が異なっても、must リンクの推移性が成立するので、 \mathbf{x}_i と \mathbf{x}_k の間にも must リンクがあると考えられる。よって、絶対クラスタリング問題の場合に、完全教師あり型のサンプルが与えられたときは、与えられた対象集合の和集合 $X = X_1 \cup \dots \cup X_N$ を作り、must リンクの推移的閉包を求めることで、より大きな教師情報 Y が生成できる。この変換により、予測精度の向上が期待できるので、絶対クラスタリング問題は準教師ありクラスタリングとして解くべきである。ここで、相対クラスタリングの場合でも must リンクの推移性は利用できるが、それは同じ Y_i にある must リンクにしか成立しないことに注意されたい。

まとめると、絶対クラスタリング問題であれば準教師ありクラスタリングとして、そうでなければ完全教師ありクラスタリングとして定式化すべきである。

5.2 対象集合の記述方法

クラスタリングをするためには、対象集合を特徴ベクトルや類似度行列などで記述する必要がある。よって、相対/絶対クラスタリング問題を解くには、対象集合についてどのような情報の記述が必要になるかを考察する。

まず、絶対クラスタリングの場合には、分割を得る操作は、未知の絶対分割のクラスタと対象との関連が分かればよい。よって、個々の対象の特徴だけに基づいて、どのクラスタに分類できるかを本質的に判別できる。3.1 節の reference matching 問題では、ある参照文字列だけ

を見れば、その他の参照文字列がなくても、隠れたクラスタ、すなわち、ある文献の実体と対応付けることができるはずである。よって、個々の対象の特徴を特徴ベクトルなどで記述すればよい。

もう一方の相対クラスタリングでは、ある対象を分類するには、他の対象との相対的な関連を考慮する必要がある。3.2 節の名詞句の coreference 問題の亀の例題では、「この亀」という指示代名詞句が、どの名詞句をさすのかということを決める必要がある。そのためには、「この亀」と他の名詞句について、互いのどちらが前にあるか、どれくらい離れているか、人を「これ」で受けるような文法違反はないかといった特徴を考慮して分類すべきである。よって、少なくとも名詞句の対の間の関係の特徴が、ベクトルやグラフといった形式で記述されていなくてはならない。

まとめると、絶対/相対クラスタリングでは、対象集合の記述に対する要求が次のように異なる。絶対クラスタリングでは、個々の対象の特徴をベクトルなどで記述すれば十分である。一方、相対クラスタリングでは、対象間の関係などの広範囲な記述が必要である。

6 まとめ

本論文では、文献 [18] で提唱した絶対/相対クラスタリングの概念と、教師ありクラスタリングとの関連について考察を深めた。

参考文献

- [1] C. C. Aggarwal, S. C. Gates, and P. S. Yu. On the merits of building categorization systems by supervised clustering. In *Proc. of The 5th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 352–356, 1999.
- [2] F. R. Bach and M. I. Jordan. Learning spectral clustering. In *Advances in Neural Information Processing Systems 16*, pp. 305–312, 2004.
- [3] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning distance functions using equivalence relations. In *Proc. of The 20th Int'l Conf. on Machine Learning*, pp. 11–18, 2003.
- [4] S. Basu, A. Banerjee, and R. Mooney. Semi-supervised clustering by seeding. In *Proc. of The 19th Int'l Conf. on Machine Learning*, pp. 19–26, 2002.
- [5] S. Basu, M. Bilenko, and R. J. Mooney. A probabilistic framework for semi-supervised clustering. In *Proc. of The 10th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 59–68, 2004.
- [6] M. Bilenko, S. Basu, and R. J. Mooney. Integrating constraints and metric learning in semi-supervised clustering. In *Proc. of The 21st Int'l Conf. on Machine Learning*, pp. 81–88, 2004.
- [7] H. Chang and D.-Y. Yeung. Locally linear metric adaptation for semi-supervised clustering. In *Proc. of*

- The 21st Int'l Conf. on Machine Learning*, pp. 153–160, 2004.
- [8] O. Chapelle, V. Vapnik, and J. Weston. Transductive inference for estimating values of functions. In *Advances in Neural Information Processing Systems 12*, pp. 421–427, 2000.
- [9] D. Cohn, R. Caruana, and A. McCallum. Semi-supervised clustering with user feedback. Technical Report TR2003–1892, Cornell University, 2003.
- [10] H. Daumé III and D. Marcu. A Bayesian model for supervised clustering with the dirichlet process prior. *Journal of Machine Learning Research*, Vol. 6, pp. 1551–1577, 2005.
- [11] T. De Bie, M. Momma, and N. Cristianini. Efficiently learning the metric with side-information. In *Proc. of the 14th Int'l Conf. on Algorithmic Learning Theory*, pp. 175–189, 2003.
- [12] R. Dubes and A. K. Jain. Validity studies in clustering methodologies. *Pattern Recognition*, Vol. 11, pp. 235–254, 1979.
- [13] V. Estivill-Castro. Why so many clustering algorithms — a position paper. *SIGKDD Explorations*, Vol. 4, No. 1, pp. 65–75, 2002.
- [14] V. Estivill-Castro and I. Lee. AUTOCLUST+: Automatic clustering of point-data sets in the presence of obstacles. In *Proc. of the 1st Int'l Workshop on Temporal, Spatial, and Spatio-Temporal Data Mining*, pp. 133–146, 2001. [LNAI 2007].
- [15] A. Ferligoj and V. Batagelj. Clustering with relational constraint. *Psychometrika*, Vol. 47, No. 4, pp. 413–426, 1982.
- [16] T. Finley and T. Joachims. Supervised clustering with support vector machines. In *Proc. of The 22nd Int'l Conf. on Machine Learning*, pp. 217–224, 2005.
- [17] M. Halkidi, D. Gunopulos, N. Kumar, M. Vazirgianis, and C. Domeniconi. A framework for semi-supervised learning based on subjective and objective clustering criteria. In *Proc. of The 5th IEEE Int'l Conf. on Data Mining*, pp. 637–640, 2005.
- [18] 神島敏弘. 絶対クラスターリングと相対クラスターリング. 人工知能学会全国大会 (第 20 回) 論文集, 2A1-1, 2006.
- [19] T. Kamishima and F. Motoyoshi. Learning from cluster examples. *Machine Learning*, Vol. 53, pp. 199–233, 2003.
- [20] 神島敏弘, 赤穂昭太郎, 元吉文男. クラスター例からの学習 — クラスター属性の利用. 人工知能学会論文誌, Vol. 18, No. 2, pp. 86–95, 2003.
- [21] 神島敏弘, 美濃導彦, 池田克夫. 帰納学習を用いた図面部品の抽出と分類のための規則の形成. 情報処理学会論文誌, Vol. 36, No. 3, pp. 614–626, 1995.
- [22] S. D. Kamvar, D. Klein, and C. D. Manning. Spectral learning. In *Proc. of the 18th Int'l Joint Conf. on Artificial Intelligence*, pp. 561–566, 2003.
- [23] D. Klein, S. D. Kamvar, and C. D. Manning. From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. In *Proc. of The 19th Int'l Conf. on Machine Learning*, pp. 307–314, 2002.
- [24] B. Kulis, S. Basu, I. Dhillon, and R. Mooney. Semi-supervised graph clustering: A kernel approach. In *Proc. of The 22nd Int'l Conf. on Machine Learning*, pp. 457–464, 2005.
- [25] N. Kumar, K. Kummamuru, and D. Paranjpe. Semi-supervised clustering with metric learning using relative comparisons. In *Proc. of The 5th IEEE Int'l Conf. on Data Mining*, pp. 693–696, 2005.
- [26] X. Li and D. Roth. Discriminative training of clustering functions: Theory and experiments with entity identification. In *Proc. of the 9th Conf. on Computational Natural Language Learning*, pp. 64–71, 2005.
- [27] Z. Lu and T. K. Leen. Semi-supervised learning with penalized probabilistic clustering. In *Advances in Neural Information Processing Systems 17*, pp. 849–856, 2005.
- [28] A. McCallum and B. Wellner. Conditional models of identity uncertainty with application to noun coreference. In *Advances in Neural Information Processing Systems 17*, pp. 905–912, 2005.
- [29] D. Mochihashi, G. Kikui, and K. Kita. Learning non-structural distance metric by minimum cluster distortions. In *Proc. of the Empirical Methods in Natural Language Processing (EMNLP-2004)*, pp. 341–348, 2004.
- [30] 小山聡, 田中克己. オブジェクト識別のための人間の教師を用いない距離学習手法. 電子情報通信学会技術研究報告, AI 2006–8, 2006.
- [31] M. Schultz and T. Joachims. Learning a distance metric from relative comparisons. In *Advances in Neural Information Processing Systems 16*, pp. 41–48, 2004.
- [32] N. Tishby, F. C. Pereira, and W. Bialek. The information bottleneck method. In *Proc. of The 37th Annual Allerton Conference on Communications, Control and Computing*, 1999.
- [33] A. K. H. Tung, J. Hou, and J. Han. Spatial clustering in the presence of obstacles. In *Proc. of The 17th Int'l Conf. on Data Engineering*, pp. 359–367, 2001.
- [34] 上野孝治, 飯田龍, 乾健太郎, 松本裕治. 半教師ありクラスターリングによる動詞辞書への大規模用例付与. 言語処理学会第 12 回年次大会論文集, pp. 1123–1126, 2006.
- [35] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl. Constrained K-means clustering with background knowledge. In *Proc. of The 18th Int'l Conf. on Machine Learning*, pp. 577–584, 2001.
- [36] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. In *Advances in Neural Information Processing Systems 15*, pp. 521–528, 2003.
- [37] K. Y. Yip, D. W. Cheung, and M. K. Ng. On discovery of extremely low-dimensional clusters using semi-supervised projected clustering. In *Proc. of The 21st Int'l Conf. on Data Engineering*, pp. 329–340, 2005.
- [38] S. X. Yu and J. Shi. Segmentation given partial grouping constraints. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 2, pp. 173–183, 2004.