

絶対クラスタリングと相対クラスタリング

Absolute Clustering and Relative Clustering

神島 敏弘^{*1}

Toshihiro Kamishima

^{*1}産業技術総合研究所

National Institute of Advanced Industrial Science and Technology (AIST)

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

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。文献 [Everitt 93] では、クラスタを「内的結合 (internal cohesion) と外的分離 (external isolation) が達成されるような部分集合」と定義を試みてはいるが、形式的な定義な困難であるとも同時に述べている。すなわち、クラスタリングは well-defined な問題ではない。このことは、文献 [林 79, Dubes 79, Estivill-Castro 02] などを含め、しばしば指摘されている。

そこで、外的規準を用いた妥当性の評価が行われてきた。これは、妥当だということがわかっている分割の例題を準備し、獲得されたクラスタが例題と類似していれば妥当とするものである。ここで発想を転換して「こうした例題があるのなら、これらの情報を使ってクラスタリングをすれば良いのではないか?」というアイデアに基づくのが、教師ありクラスタリングである。本論文では、この教師ありクラスタリングにおいて、公理的に区別すべき、絶対クラスタリングと相対クラスタリングの 2 種類の問題があることを述べる。

2. 節では教師ありクラスタリングについて、3. と 4. 節では絶対/相対クラスタリング問題と具体例について、そして 5. と 6. 節では議論とまとめを述べる。

2. 教師ありクラスタリング

教師ありクラスタリング問題の定義や呼称は、個々の研究者がそ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、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 [Chang 04] 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リングとする。逆に、文献 [Basu 02] のように、ラベルあり/なし混在データをクラスタ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、制約付きクラスタリング (constrained clustering) と教師ありクラスタリングを区別する。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リング、でなければ教師ありクラスタリングとする。Wagstaff らの COP-KMEANS[Wagstaff 01] を例に説明する。この方法では、cannot リンクと must リンクの二種類の制約条件がある。cannot リンクで結ばれた対象の対は同じクラスタの要素となつてはならず、must リンクで結ばれた対象の対は必ず同じ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、制約のある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、代表的な準教師ありクラスタリングとして紹介されることもある方法だが、ここでは制約付きクラスタリングに分類する。制約付きクラスタリングには、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 [Ferligoj 82] や、障害物といった制約 [Tung 01, Estivill-Castro 01, Zaiane 02] などがある。

上記のことから、ラベル集合が事前には未知であるような分類問題を、教師情報を一般化して解決する問題を、広義の教師ありクラスタリング (supervised clustering) とする。この、広義の教師ありクラスタリングをさらに、(完全) 教師ありクラスタリング ((fully) supervised clustering) と準教師ありクラスタリング (semi-supervised clustering) に分類する^{*2}。完全教師ありと準教師ありクラスタリングは、教師情報の与え方によって区別する。完全教師ありの場合では、適切なクラスタの具体例や、must/cannot リンク情報などを与えた対象集合を訓練事例とし、この訓練事例の集合を訓練サンプルとする。一方、準教師ありクラスタリングでは、一つの対象集合に適切な分割や must/cannot リンクなどの情報を附加したものが与えられる。完全教師ありクラスタリングの研究としては [神島 95, 神島 03b, Kamishima 03a, Daumé III 05, Finley 05] などがある。

準教師ありでの教師情報としては、Wagstaff の must/cannot リンク制約を使ったものが多い [Klein 02, Cohn 03, Xing 03, Chang 04, Basu 04, McCallum 05, Yu 04, Bilenko 04]。また、同じクラスタになるべき対象の部分集合を与える方法 [Bar-Hillel 03, Bach 04] 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。また、ラベル情報を部分的に与える準教師ありクラスタリングも考えられるが、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。他に、Tishby らの情報ボトルネック [Tishby 99, Slonim 00] など、クラスタリングされ

^{*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 [Klein 02]

^{*2} 以後、単に教師ありクラスタリングといった場合は、広義のそれをさすものとする。

る対象と関連がある副情報を教師情報とするものなどもある。しかし、ここでは must/cannot リンク制約を使うものを中心に考える。

さらに、準教師ありクラスタリングの下位概念として、学習時に与えられた対象集合を分割することだけが目的で、新たな対象の分類はできなくてもよいという問題がある。これは、リッジ回帰を対象にした transductive 学習 [Chapelle 00] と類似しているので transductive クラスタリングと呼ぶ。

整理すると、クラス分類の下位概念に準教師ありクラス分類がある。一方、クラスタリングには、制約付きクラスタリングと広義の教師ありクラスタリングがあり、広義の教師ありクラスタリングは準教師ありと完全教師ありに分けられる。さらに、transductive クラスタリングは準教師ありクラスタリングの部分概念である。最後に、文献 [Chang 04] の提案を採用し準教師ありクラスタ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ことにする。

教師ありクラスタリングを解くアプローチは 2 種類ある。一つは、制約ができるだけ満たすように対象間の類似度関数を学習し、その関数で求めた類似度を、既存のクラスタリング手法へ入力してクラスタを得る方法である。もう一つは、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導入し、これを最適化するクラスタを探索する方法である。ここでは、前者をラッパー型、後者を統合型と呼ぶことにする。

学習の目標も多様である。与えられた事例集合の妥当なクラスタを得ることだけが目標であるものや、妥当なクラスタを得たあと、新たな対象を適切に分類できることも目標であるものもある。さらに、任意の対象集合の適切なクラスタを発見できる規則や関数を学習するものもある。

3. 絶対クラスタリングと相対クラスタリング

前節では、教師ありクラスタリングの概観を述べた。この教師ありクラスタリングとの関連は後に述べるとして、先に絶対クラスタリングと相対クラスタリングの定義を与える。まず、クラスタリング関数の形式的定義を示す。全ての可能な対象の集合 \mathcal{X} の任意の部分集合を $X = \{x_i\}_i^N$ で表す。 X の分割とは、 X について網羅的で、互いに素な X の部分集合の集合 $\{C_j\}_j^K$ であり、 X の可能な分割の全体の集合を $\Pi(X)$ とする。この部分集合 C_j をクラスタと呼ぶ。与えられた X に対して、クラスタリング関数 $\pi(X)$ は $\Pi(X)$ 中の適切な分割を出力する。関数 $\text{isc}(x_i, x_j, \pi(X))$ は、 $i \neq j$ なる対象 x_i と x_j が、 $\pi(X)$ 中の同一クラスタの要素、すなわち、 $x_i \in C_k \wedge x_j \in C_k$, $\exists k \in \{1, \dots, K\}$ 、であるなら 1 をとり、そうでなければ 0 をとる関数である。ここで、クラスタリング関数 $\pi(X)$ が次の条件を満たす場合絶対クラスタリング (absolute clustering) といい、そうでない場合に相対クラスタリング (relative clustering) という：

$$\begin{aligned} \text{isc}(x_i, x_j, \pi(X_1)) &= \text{isc}(x_i, x_j, \pi(X_2)), \\ \forall x_i, \forall x_j \in X_1 \cap X_2, x_i \neq x_j, \forall X_1, \forall X_2 \subseteq \mathcal{X} \end{aligned} \quad (1)$$

直観的には、対象 x_i と x_j が同じクラスタに分類されるかどうかが、クラスタリングする対象集合とは独立に決まる、すなわち、 x_i と x_j のみに依存してい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は、観測されない絶対的に不变な \mathcal{X} の分割である絶対分割 $\pi^*(\mathcal{X})$ があり、その絶対分割中で同じクラスタの要素が同じクラスタに分類される。一方、相対クラスタリングでは絶対分割はない。

この絶対/相対クラスタリングは、教師あり順序付け問題での、絶対/相対順序付け [神島 05b, Kamishima 05a] と対応付けて考えることができる。

4. 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。

4.1 Reference Matching, Record Linkage, Identity Uncertainty

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。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「神島敏弘」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。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。より一般的な問題として、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。

この問題には、文字列間の類似度を編集距離などで測り、既存のクラスタリング手法を適用して処理してきた。しかし、具体例が作成できるため、教師ありクラスタリングが最近はこの問題に適用されている。同じ文献や人物を must リンクで、違うものを cannot リンクで表した教師情報を用いる。

この問題を、クラスタリング関数としてみたとき、式 (1) の条件を満たすべきかを考察する。文献の参照を示す文字列 A と B が、真に同じ文献を示すものであれば、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されるべきである。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。

4.2 名詞句の coreference

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。例えば、総理大臣の動向の記事を考える。その中で「小泉氏」「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。この問題も must/cannot リンクを用いて教師ありクラスタリング手法が適用できる。

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 から成る文書を考える。

- A. 親亀がいた。
- B. この亀に子亀を乗せた。
- C. この亀に孫亀を乗せた。

coreference 問題を考えると、生成されるべきクラスタは {親亀, (文 B の) この亀}, {子亀, (文 C の) この亀}, そして {孫亀} である。この場合は「親亀」と「(文 C の) この亀」は異なるクラスタの要素である。ここで、文 B を取り去ると「(文 B の) この亀」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, (文 C の) この亀} 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「(文 C の) この亀」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．

4.3 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．この問題を解くクラスタリング関数を獲得することを想定する．この問題は，二つの画素の間にある別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ため，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，未知だが不变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．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，今週の政治と経済の記事で構成されているなら，カテゴリによって分類するのが目標とする．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．

DNA から転写された RNA 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/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．

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，4.2 節の問題では，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，部分的な教師情報を付加された訓練集合を分割することで，要素の追加や削除は考慮しない．すなわち，与えられた集合が \mathcal{X} に相当し，transductive アルゴリズム自体が関数となって $\pi(\mathcal{X})$ を導くことが目的である．以上の考察から，絶対/相対クラスタリングの区別は，transductive クラスタリングの場合には考慮する必要がない．

よって，以降は non-transductive な準教師ありクラスタリング^{*3}と完全教師ありクラスタリングについて議論する．絶対/相対クラスタリングの違いは，(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影響すると考える．以下，これらについて述べる．

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与え方が異なる．完全教師ありの場合は，対象集合 X に，それに対する適切な分割の情報 $\pi^*(X)$ を付加したものを，複数個集めた集合が訓練サンプル集合になる．一方，準教師ありの場合では，一つの対象集合に must/cannot リンクのような分割の部分情報を付加して，訓練サンプルとする．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

^{*3} 以降は non-transductive なものを単に準教師ありクラスタリングと呼ぶ

対の対象が同じクラスタになるかどうかが，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．各対象の所属は，それらの対象を含む対象集合に依存するため，相対クラスタリング問題では，完全教師ありのアプローチをしなければならない．もし，準教師アプローチをとると，その集合に一つ要素を加えただけで，教師情報として与えた must/cannot リンク制約が成立しなくなるかもしれないからである．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，対象集合集合 X_1 と X_2 があり， X_1 中の x_i, x_j 間と X_2 中の x_j, x_k 間に must リンク制約があったとする．式 (1) の条件から，対象集合が異なっていても，must リンクの推移性が成立するので， x_i と 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で，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 $X_1 \cup X_2 \cup \dots$ を作り，must リンクの推移的閉包を求めて，より多くの教師情報を得るべきである．

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アプローチを採用し，そうでなければ完全教師ありアプローチにすべきである．

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，対象集合をベクトルや行列などの形式で記述する必要がある．教師なしクラスタリングでは，対象を属性ベクトルか，非類似度行列を用いて，対象集合を記述することが多い．前者は，対象の特徴を記述する数値型やカテゴリ型の変数のベクトルを考え，個々の対象を実現値を与えたベクトルで表す．後者は，対象 x_i と x_j 間の非類似度を， i 行 j 列目の要素とする行列である．一般には，対称行列で要素は非負である．

ここで，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は，分割を得る操作は，隠れた絶対分割のクラスタと対象との関連が分かれよい．よって，対象の個体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．4.1 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記述するベクトルなどを採用すればよい．

一方，相対クラスタリング問題ではどうであろうか？相対クラスタリング問題では，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．4.2 節の 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解決する必要がある．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，最低で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もので必ず記述されなくてはならない．文献 [神島 03b, Kamishima 03a] では，さらに集合全体やクラスタの特徴といったものの記述も行っている．

まとめると，対象集合の記述において絶対/相対クラスタリングでは，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，対象間の関係などをより広範囲に記述できなくてはならない．

6. まとめ

本論文では、近年、研究が始まった教師ありクラスタリングの研究を俯瞰し、その組織的な分類を試みた。さらに、クラスタリング問題は、公理的な相違がある、絶対クラスタリングと相対クラスタリングに分類することができることを述べた。絶対クラスタリング問題は、準教師ありクラスタリングのアプローチで解くべきこと、また、個々の対象をベクトルなどで記述すれば十分であることを示した。一方、相対クラスタリング問題には、完全教師ありクラスタリングアプローチを採用すべきであること、また、個々の対象だけではなく、対象間の関係の記述も必要になることを示した。

クラスタリングは問題の粒度 (granularity) を変えることのできる唯一の方法と考える。だが、well-defined ではないという根本的な問題があった。教師ありクラスタリングは、クラスタリング問題を well-defined にする有望なアプローチと考えいる。よって、効果的なアルゴリズムが考案されれば、その影響は非常に大きいだろう。教師ありクラスタリングを解くアプローチには、ポテンシャル関数を定義する方法や、距離関数を学習する方法など幾つかの方法がある。今後は、これらのアプローチと絶対/相対クラスタリングとの関連について考察を深めたい。

参考文献

- [Bach 04] Bach, F. R. and Jordan, M. I.: Learning Spectral Clustering, in *Advances in Neural Information Processing Systems 16*, pp. 305–312 (2004)
- [Bar-Hillel 03] Bar-Hillel, A., Hertz, T., Shental, N., and Weinshall, D.: Learning Distance Functions using Equivalence Relations, in *Proc. of The 20th Int'l Conf. on Machine Learning*, pp. 11–18 (2003)
- [Basu 02] Basu, S., Banerjee, A., and Mooney, R.: Semi-supervised Clustering by Seeding, in *Proc. of The 19th Int'l Conf. on Machine Learning*, pp. 19–26 (2002)
- [Basu 04] Basu, S., Bilenko, M., and Mooney, R. J.: 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)
- [Bilenko 04] Bilenko, M., Basu, S., and Mooney, R. J.: Integrating Constraints and Metric Learning in Semi-Supervised Clustering, in *Proc. of The 21st Int'l Conf. on Machine Learning*, pp. 81–88 (2004)
- [Chang 04] Chang, H. and Yeung, D.-Y.: Locally Linear Metric Adaptation for Semi-Supervised Clustering, in *Proc. of The 21st Int'l Conf. on Machine Learning*, pp. 153–160 (2004)
- [Chapelle 00] Chapelle, O., Vapnik, V., and Weston, J.: Transductive Inference for Estimating Values of Functions, in *Advances in Neural Information Processing Systems 12*, pp. 421–427 (2000)
- [Cohn 03] Cohn, D., Caruana, R., and McCallum, A.: Semi-supervised Clustering with User Feedback, Technical Report TR2003-1892, Cornell University (2003)
- [Daumé III 05] Daumé III, H. and Marcu, D.: A Bayesian Model for Supervised Clustering with the Dirichlet Process Prior, *Journal of Machine Learning Research*, Vol. 6, pp. 1551–1577 (2005)
- [Dubes 79] Dubes, R. and Jain, A. K.: Validity Studies in Clustering Methodologies, *Pattern Recognition*, Vol. 11, pp. 235–254 (1979)
- [Estivill-Castro 01] Estivill-Castro, V. and Lee, I.: 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]
- [Estivill-Castro 02] Estivill-Castro, V.: Why So Many Clustering Algorithms — A Position Paper, *SIGKDD Explorations*, Vol. 4, No. 1, pp. 65–75 (2002)
- [Everitt 93] Everitt, B. S.: *Cluster Analysis*, Edward Arnold, third edition (1993)
- [Ferligoj 82] Ferligoj, A. and Batagelj, V.: Clustering with Relational Constraint, *Psychometrika*, Vol. 47, No. 4, pp. 413–426 (1982)
- [Finley 05] Finley, T. and Joachims, T.: Supervised Clustering with Support Vector Machines, in *Proc. of The 22nd Int'l Conf. on Machine Learning*, pp. 217–224 (2005)
- [林 79] 林 知己夫：クラスター分析の意味, 数理科学, No. 190, pp. 5–8 (1979)
- [神嶌 95] 神嶌 敏弘, 美濃 審彦, 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)
- [Kamishima 03a] Kamishima, T. and Motoyoshi, F.: Learning from Cluster Examples, *Machine Learning*, Vol. 53, pp. 199–233 (2003)
- [神嶌 03b] 神嶌 敏弘, 赤穂 昭太郎, 元吉 文男：クラスタ例からの学習 — クラスタ属性の利用, 人工知能学会論文誌, Vol. 18, No. 2, pp. 86–95 (2003)
- [Kamishima 05a] Kamishima, T., Kazawa, H., and Akaho, S.: Supervised Ordering — An Empirical Survey, in *Proc. of The 5th IEEE Int'l Conf. on Data Mining*, pp. 673–676 (2005)
- [神嶌 05b] 神嶌 敏弘, 賀沢 秀人, 赤穂 昭太郎：教師あり順序付け — 手法の比較実験, 2005 年情報論的学習理論ワークショップ, pp. 213–218 (2005)
- [Klein 02] Klein, D., Kamvar, S. D., and Manning, C. D.: 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)
- [McCallum 05] McCallum, A. and Wellner, B.: Conditional Models of Identity Uncertainty with Application to Noun Coreference, in *Advances in Neural Information Processing Systems 17*, pp. 905–912 (2005)
- [Slonim 00] Slonim, N. and Tishby, N.: Agglomerative Information Bottleneck, in *Advances in Neural Information Processing Systems 12*, pp. 617–623 (2000)
- [Tishby 99] Tishby, N., Pereira, F. C., and Bialek, W.: The Information Bottleneck Method, in *Proc. of The 37th Annual Allerton Conference on Communications, Control and Computing* (1999)
- [Tung 01] Tung, A. K. H., Hou, J., and Han, J.: Spatial Clustering in the Presence of Obstacles, in *Proc. of The 17th Int'l Conf. on Data Engineering*, pp. 359–367 (2001)
- [Wagstaff 01] Wagstaff, K., Cardie, C., Rogers, S., and Schroedl, S.: Constrained K-means Clustering with Background Knowledge, in *Proc. of The 18th Int'l Conf. on Machine Learning*, pp. 577–584 (2001)
- [Xing 03] Xing, E. P., Ng, A. Y., Jordan, M. I., and Russell, S.: Distance Metric Learning, with Application to Clustering with Side-Information, in *Advances in Neural Information Processing Systems 15*, pp. 521–528 (2003)
- [Yu 04] Yu, S. X. and Shi, J.: Segmentation Given Partial Grouping Constraints, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 2, pp. 173–183 (2004)
- [Zaijane 02] Zaijane, O. R. and Lee, C.-H.: Clustering Spatial Data when Facing Physical Constraints, in *Proc. of The 2nd IEEE Int'l Conf. on Data Mining*, pp. 737–740 (2002)