

語彙統語パターンにもとづく制約付き分布クラスタリング

Constrained Distributional Clustering of Words using Lexico-syntactic Patterns

鍛治伸裕^{1*} 喜連川優¹
Nobuhiro Kaji¹ and Masaru Kitsuregawa¹

¹ 東京大学生産技術研究所
¹ Institute of Industrial Science, University of Tokyo

Abstract: This paper proposes to incorporate supervision in the form of pair-wise constraints into word clustering. Following the framework of constrained clustering, or semi-supervised clustering, we extended distributional clustering algorithm by using the Markov random field as a prior distribution on hidden variables. In addition, we propose to automatically induce constraints from corpora using lexico-syntactic patterns, while constraints are usually given by hand in literatures. Our experimental results demonstrated that the quality of clustering was significantly improved by constraints.

1 はじめに

クラスタリングは伝統的に教師なし学習の問題として定式化されてきた。しかし、近年では、部分的に教師データを利用する半教師ありクラスタリングという枠組みが提案されている [3, 4, 5, 15, 17, 20, 25]。この枠組みでは、同じ (または異なる) クラスタに属すべきデータの組が制約として与えられ、クラスタリングのモデルに組み込まれる。半教師ありクラスタリングは、教師データが制約という形で与えられることから、制約付きクラスタリングとも呼ばれる。

クラスタリングは、自然言語処理や情報検索の多くの場面で利用される。例えば単語のクラスタリングはその一例である [18, 22]。単語をクラスタリングして意味的に類似した語をまとめあげるとは、自然言語処理における基本技術の 1 つであり、言語モデル構築 [7, 9]、テキスト分類 [2]、項構造解析 [13] などに応用されている。

これまで単語のクラスタリングは教師なし学習の問題として扱われてきた。しかし、単語のクラスタリングにおける制約 (=教師データ) の有効性は明らかである。例えば、シソーラスなどに登録されている類義語は、単語のクラスタリングのための制約と考えることができる。このような制約をうまく利用できれば、クラスタリングの精度向上が期待できる。また、制約を使ってクラスタリング結果を制御することができれば、

個別のアプリケーションごとに異なる制約を作りこんでおき、アプリケーションに応じてクラスタリングの仕方を変えることも可能になる。

本論文では、制約付きクラスタリングの枠組みを用いて、単語のクラスタリングに制約を導入することを提案する。クラスタリングには、分布クラスタリング [22] を拡張した確率モデルを用いた。拡張されたモデルは、分布クラスタリングにおける隠れ変数の事前分布にマルコフ確率場を導入することによって、制約を破る隠れ変数にペナルティを与えたものになっている [4]。モデルのパラメータは、平均場近似を用いた EM アルゴリズムで推定した [15]。

一般的な制約付きクラスタリングでは、制約は人手で与えられることが前提となっている。しかし、本論文では、語彙統語パターンを用いてコーパスから類義語を自動獲得して、それにもとづいて制約を導出する方法を提案する。前述のように、シソーラスを制約として使うことも考えられるが、コーパスから制約を自動的に導出することによって、新語や固有表現など既存のシソーラスには登録されていない語も制約に加えることが可能になる。

クラスタリングの結果は、人手で作られたシソーラスを用いて評価した。同じクラスタに所属しやすい単語ペアを無作為に取り出し、シソーラスに基づく類似度 [14, 23] を求めたところ、自動的に導出された制約を与えることによって類似度が向上することが確認できた。この結果から、提案モデルのクラスタリング結果は、制約を用いない場合よりも人間の直感に近いこ

*連絡先: 東京大学生産技術研究所 喜連川研究室
〒153-8505 東京都目黒区駒場 4-6-1
E-mail: kaji@tkl.iis.u-tokyo.ac.jp

とが分かった。

2 分布クラスタリングと語彙統語パターン

まず、提案手法の基礎である分布クラスタリングと語彙統語パターンを簡単に説明する。

2.1 分布クラスタリング

分布クラスタリングは、隠れ変数モデルにもとづいてソフトクラスタリングを行う手法である。単語をクラスタリングする場合は、名詞 n と動詞 v の係り受けの出現確率が次のように定義される¹[22]。

$$p(n, v) = \sum_z p(n, v, z) = \sum_z p(n|z)p(v|z)p(z) \quad (1)$$

ここで z は隠れ変数であり、名詞と動詞の組に関する潜在的な意味カテゴリであると解釈することができる。モデルのパラメータ $p(n|z)$ と $p(v|z)$ は、語のカテゴリへの確率的な所属を表現しており、これがソフトクラスタリングを行っていることに相当する。

N 個の観測データ $D = \{(n_i, v_i)\}_{i=1}^N$ が与えられた時、各観測データに対応する隠れ変数を $Z = \{z_i\}_{i=1}^N$ とすると、観測データ D が生成される確率 $p(D)$ は次のように定義される。

$$p(D) = \sum_Z p(D|Z)p(Z) \quad (2)$$

$$p(D|Z) = \prod_{i=1}^N p(n_i, v_i|z_i) \quad (3)$$

$$= \prod_{i=1}^N p(n_i|z_i)p(v_i|z_i) \quad (4)$$

$$p(Z) = \prod_{i=1}^N p(z_i) \quad (5)$$

式 (2) の \sum_Z は、 N 個の隠れ変数 $Z = (z_1, z_2, \dots, z_N)$ の値の全組み合わせに対して和をとることを意味する。 $P(D|Z)$ は観測データの条件付き確率、 $P(Z)$ は隠れ変数の事前確率である。

モデルのパラメータは EM アルゴリズムなどを用いて推定することができる [12, 24, 27]。E ステップでは、現在のパラメータにもとづいて $p(z_i|D)$ を計算す

る。 $p(z_i|D)$ は、隠れ変数が互いに独立であるため、下の式を使って効率的に計算できる。

$$p(z_i|D) = \frac{p(n_i, v_i, z_i)}{\sum_z p(n_i, v_i, z)} \quad (6)$$

M ステップでは、E ステップの結果をもとに最尤推定を行って、パラメータの値を更新する。そして、この処理をパラメータが収束するまで続ける

2.2 語彙統語パターン

語彙統語パターンとは、上位下位関係など、特定の意味関係にある単語ペアを獲得するための抽出パターンのことである。例えば、上位下位関係にある単語を抽出するためには「X such as Y」などのパターンがよく用いられる [10]。パターンの X と Y にマッチする単語をコーパスから収集することにより、上位下位関係にある単語を自動獲得することができる。

語彙統語パターンは類義語の獲得にも適用できる [6, 8]。例えば Bollegala らは、自動学習した語彙統語パターンを使って類義語の獲得を行っており、「X/Y」や「X or Y」などが有効であったと報告している。彼らの実験は英語を対象としているが、日本語でも同様に「X や Y」などが有効であると考えられる。

3 制約付き分布クラスタリング

語彙統語パターンを使うと、例えば日本酒とブランデーが類義語であることを重み (または確信度) 付きで学習することができる。さらに、その結果として「日本酒を飲む」と「ブランデーを飲む」などが意味的に類似していることも分かる。これを分布クラスタリングの枠組みで考えると、ある (n, v) と (n', v') の隠れ変数は同じ値をとる、という制約が重み $w (> 0)$ 付きで与えられたことに相当する。

本論文ではこのような制約を用いた分布クラスタリングを提案する。以下では、まず制約付き分布クラスタリングの確率モデルを説明する。次に、語彙統語パターンを用いて制約を自動的に導出する手法を述べる。

3.1 確率モデル

制約付き分布クラスタリングの確率モデルは、基本的に制約のない場合と同様である。唯一の違いは、隠れ変数の事前確率 $p(Z)$ である。

我々は先行研究にならぬ、隠れ変数の事前確率 $p(Z)$ にマルコフ確率場を導入することによって、制約を反映したモデルを構築する [4]。今、観測データ (n_i, v_i)

¹ 正確には v は動詞ではなく格助詞と動詞の組だが、簡単のため動詞と呼ぶ。

と (n_j, v_j) の間に制約が与えられていて、その重みが $w_{ij} (> 0)$ であるとする。このとき隠れ変数の事前分布 $p(Z)$ を以下のように定義する。

$$p(Z) = \prod_{i=1}^N p(z_i) \times \frac{\exp\{-\sum_{i,j} w_{ij} \delta(z_i \neq z_j)\}}{G}$$

ただし $\delta(\cdot)$ はデルタ関数であり、隠れ変数 z_i と z_j の値が異なる場合は 1、それ以外は 0 をとる。 G は右辺第 2 項の正規化項である。式 (5) とは異なり、隠れ変数同士は独立でない。

このモデルが完全データ (D, Z) を生成する確率の対数 $\log p(D, Z)$ を求めてみると、制約を破るような Z が出現しにくくなっていることが確認できる。さきほど示した $p(Z)$ と前節の $p(D|Z)$ を使うと、

$$\log p(D, Z) = \sum_{i=1}^N \log p(n_i|z_i) \quad (7)$$

$$+ \sum_{i=1}^N \log p(v_i|z_i) \quad (8)$$

$$+ \sum_{i=1}^N \log p(z_i) \quad (9)$$

$$- \sum_{i,j} w_{ij} \delta(z_i \neq z_j) - \log G \quad (10)$$

を得る。右辺の最初の 3 項は、制約無し分布クラスタリングの対数尤度と同じである。そして、第 4 項は、制約を破っている隠れ変数に対するペナルティに相当する。

3.2 制約の導出

通常の制約付きクラスタリングでは、制約は人手で与えられることが前提となっている。しかし、本論文では、語彙統語パターンを用いて制約を自動的に導出する。制約導出の手順としては、まず語彙統語パターンを利用して表 1 のような類義語集合をコーパスから獲得する。そして、これをもとに制約を導出する。

名詞の類義語を獲得するためには、以下の 5 種類の語彙統語パターンを用いた。

X や Y X も Y も X と Y と X、Y、 X、Y、

これらのパターンを形態素解析したコーパスに適用して、X と Y にマッチする名詞 (または複合名詞) を抽出する。動詞についても「X したり Y したり」というパターンを用いて同様の処理を行った。

しかし、単純にパターンにマッチした語を収集したのではノイズが混じってしまうため、これを除去する必要がある。そこで我々はグラフの連結度にもとづく手

法を考案した。パターンで収集された語集合はグラフとみなすことができる (語が頂点、類義関係が辺)。そして、密な辺を持つ頂点集合 (典型的にはクリーク) は信頼できる類義語集合であると考えられる。そこで、このグラフから連結度の高い頂点集合を抽出して使うことにした。

連結度の高い頂点集合は次のようなアルゴリズムで求めた。基本的な処理はボトムアップクラスタリングと同じである。まず最初に、各語が大きさ 1 のクラスタを形成しているような状態を作る。そして、適当な順番で 2 つのクラスタを選び、それらが以下の条件を満たせば併合する、という操作を繰り返す。

- 併合後のクラスタに属する語が連結グラフを形成している。
- その連結グラフにおいて、全ての頂点が残りの頂点の過半数と辺で結ばれている。

この操作を行うことにより、連結度の高い頂点集合がクラスタリングされていく。以上の処理を、併合できるクラスタが存在しなくなるまで繰り返した後、大きさが S 以上のクラスタ²を信頼できる類義語集合として取り出す (表 1)。

さて、ここまでの処理で名詞と動詞の類義語集合を得ることができた。最後に、この類義語集合を用いて制約を導出する。ここで 2 つの観測データ (n, v) と (n', v') を考える。ただし $n \neq n'$ または $v \neq v'$ とする。このとき、もし n と n' が同一の類義語集合に属しており、なおかつ v と v' も同一の類義語集合に属していれば、この 2 つのデータの間には制約を与える。さらに、 n と n' が同じ類義語集合に属しており、なおかつ v と v' が同じ動詞である場合も制約を与える。逆の場合も同様である。いずれの場合も制約の重み w は

$$w = \frac{1}{\sqrt{C(n, v) \times C(n', v')}} \quad (11)$$

とした。 $C(n, v)$ は観測データにおける (n, v) の頻度である。

表 1: 類義語集合の例

浴場	露天風呂	サウナ	ジャグジー	湯	水風呂
マスク	チーク	ファンテーション	口紅	アイライン	
メモリ	CPU	マザーボード	FDD	HDD	メモリー
揚げる	炒める	焼く	煮る	ゆでる	茹でる
感心する	驚く	感動する	共感する	納得する	

²実験では $S = 5$ とした。

4 パラメータ推定

提案モデルのパラメータは EM アルゴリズムを用いて推定する。

4.1 平均場近似を用いた E ステップの計算

制約付き分布クラスタリングにおける EM アルゴリズムの手続きは、制約の無い場合 (2 節) と同じである。しかし、制約付き分布クラスタリングでは隠れ変数が独立でないため、 $p(z_i|D)$ を効率的に計算することができない。すなわち、E ステップで以下の確率計算を行わなくてはならない。

$$p(z_i|D) = \sum_{Z_{-i}} p(Z|D) \quad (12)$$

$$\propto \sum_{Z_{-i}} p(D|Z)p(Z) \quad (13)$$

$\sum_{Z_{-i}}$ は、 z_i を除いた $N-1$ 個の隠れ変数の全組み合わせに対して和を取るという意味である。

式 (13) を求めるのに必要な計算量は、隠れ変数の種類を $|Z|$ とすると $O(|Z|^{N-1})$ であり、一般的には計算不可能である。そこで平均場近似を用いて近似計算を行う [15]。マルコフ連鎖モンテカルロ法など、他の方法で E ステップの計算を行うことも可能であるが [17, 20]、ここでは計算効率を考慮して平均場近似を採用する。

平均場近似では、確率分布 $p(Z|D)$ を以下のような形の分布 $q(Z)$ で近似する。

$$q(Z) = \prod_{i=1}^N q_i(z_i) \quad (14)$$

$q(Z)$ は、真の分布とは違って隠れ変数が独立となっている。そのため、 $p(Z|D)$ の代わりに $q(Z)$ を使えば、制約の無い場合と同じく、効率的に E ステップの計算を行うことができる。

近似分布 $q(Z)$ のパラメータは、真の分布 $p(Z|D)$ からの KL ダイバージェンスが最小となるものを選ぶ。すなわち、近似分布を導出する作業は、任意の i に対して $\sum_z q_i(z) = 1$ という条件のもとで、以下の目的関数を最小化する問題となる。

$$\sum_Z q(Z) \log \frac{q(Z)}{p(Z|D)} \quad (15)$$

ラグランジュの未定乗数を導入して、 $q_i(z)$ の極値条件を求めると、

$$q_i(z) \propto p(n_i, v_i, z) \times \exp\left\{-\sum_{j \neq i} (1 - q_j(z)) w_{ij}\right\} \quad (16)$$

を得る。

以上の議論にもとづき、近似分布のパラメータを次の更新式にもとづく反復法で求める。

$$q_i^{(t+1)}(z) \propto p(n_i, v_i, z) \times \exp\left\{-\sum_{j \neq i} (1 - q_j^{(t)}(z)) w_{ij}\right\}$$

ただし $q_i^{(t)}(z)$ は t 回目の繰り返しにおける $q_i(z)$ の値である。初期値 $q_i^{(0)}(z)$ は無作為に決定した。更新式の \sum の計算は $w_{ij} \neq 0$ となる j だけについて行えばよいので、効率的に計算することができる。

4.2 初期値の設定

EM アルゴリズムには局所最適解におちいるという問題があり、その性能は初期値に依存する。そこで、語彙統語パターンで獲得した類義語集合を用いて初期値の設定を行う。名詞 n と n' が同一の類義語集合に属する場合、確率分布 $p(c|n)$ と $p(c|n')$ が等しくなるように初期値を決めた。動詞についても同様に初期値を設定した。

5 実験

提案モデルの有効性を検証するため、制約を用いない分布クラスタリング [22] との比較実験を行った。全ての実験で隠れ変数の数は 500 とした。

5.1 データ

実験には、新聞記事 (15 年分) とウェブから収集した約 10 億文を用いた。形態素解析と構文解析にはそれぞれ Juman と KNP を用いた³。

名詞と動詞の係り受けデータは、10 億文のうち約 2 億文の構文解析結果から「が」「を」「に」「格」の係り受けを抽出して作成した。係り受けデータの総数は 61,632,587、種類数は 793,768 であった。名詞と動詞の種類数は 32,744 と 12,877 であった。

制約の導出には 10 億文の全てを用いた。語彙統語パターンを適用した後にノイズの除去を行った結果、343 の類義語集合を獲得することができた。このうち、名詞の類義語集合が 313、動詞の類義語集合が 30 であった。これに含まれる名詞の総数は 1,946、動詞の総数は 181 であった。この類義語集合を用いて制約を導出したところ、7,500,794 の係り受けデータ (76,434 種類の係り受け) に制約が与えられた。

³<http://nlp.kuee.kyoto-u.ac.jp/nl-resource/>

5.2 手順

クラスタリング結果を評価するために、同じクラスタに所属しやすい単語ペアを無作為に取り出して、それらの類似度をシソーラス [1] をもとに計算し、その平均値の比較を行った。

実験では、まず無作為に 50 個のクラスタを選び、そのそのクラスタの所属確率 $p(n|c)$ の上位 10 語を集めた。その後で、シソーラスに登録されていない語と所属確率が低い語 ($\log p(n|c) < -1$) を取り除いた。そして、同じクラスタから集められた語の全ての組み合わせに対して類似度を計算した。

類似度計算には 3 種類の方法を試した。1 つめは、2 語がシソーラス上で同じカテゴリに所属していれば 1 とし、それ以外は 0 とする類似度である。もし語義に曖昧性があって、1 語が複数のカテゴリに所属している場合は、どれか 1 つのカテゴリが一致していれば類似度を 1 とした。この類似度を EXACT-SIM と呼ぶ。残りの 2 つの類似度は、Risnik らと Kurohashi らの提案したもの [23, 14] を使った。以下ではこれらを RESNIK-SIM, KUROHASHI-SIM と呼ぶ。なお、EXACT-SIM と KUROHASHI-SIM は類似度の最大値が 1 に正規化されているが、RESNIK-SIM は正規化されていない。

5.3 結果

表 2 に実験の結果を示す。EM アルゴリズムの初期値の影響を考慮して 5 回の実験の平均値を載せている。

類似度計算法	制約なし	制約あり
EXACT-SIM	0.116	0.157
RESNIK-SIM	2.267	2.546
KUROHASHI-SIM	0.484	0.507

いずれの類似度計算を採用した場合でも、制約を与えることによって、平均類似度が向上している。このことから、提案モデルのクラスタリング結果は、制約なしの場合に比べて、人手で作成したシソーラスの分類に近いことが分かる。

6 議論

一般的に制約付きクラスタリングの枠組みでは、2 種類の制約が議論の対象となっている。1 つは、ある 2 つのデータが同じクラスタに所属するという制約で must-link と呼ばれる。もう 1 つは、2 つのデータが別のクラスタに所属するという制約で cannot-link と呼ばれる。

本論文では、must-link に相当する制約しか議論の対象にしていない。そのため、cannot-link を利用することによって、クラスタリングの精度を更に向上させることができると考えている。モデル自体は cannot-link を加えた形に容易に拡張できるが、問題となるのは cannot-link の制約をどのように用意するかである。1 つの方法は人手で作成することであるが、cannot-link を自動的に導出するという方向も検討していきたい。

今後の課題として、あるアプリケーションに対して制約を作りこみ、アプリケーションに特化したクラスタリングを行うことが挙げられる。クラスタリングというタスクには絶対的な正解がなく、何をもって正しいクラスタリングとするかはアプリケーションに依存するところが大きい。このことから、アプリケーションに特化したクラスタリングを実現することは重要であると考えている。

7 関連研究

Takamura らはスピンモデルを用いて単語の評価極性を抽出する手法を提案した [26]。Takamura らのモデルは、提案モデルの $p(Z)$ とほぼ同じ形をしており、2 つのモデルには類似性が見られる。さらに、解いている問題も類似している。単語の評価極性の抽出は、好評/不評/中立への分類問題であり、これはある種のクラスタリングであると解釈することができる。以上の考察にもとづくと、提案モデルは単語の評価極性抽出にも適用できると思われる。今後は、実際に有効性を検証することも考えている。

3 節で説明したグラフの連結度にもとづくノイズ除去の手法は、Matsuo らが提案したクラスタリングアルゴリズムと類似している [18]。しかし、我々は語をクラスタリングするだけでなく、辺が密になっていない頂点を捨てることによって、ノイズの除去を同時に行っている点が異なる。

本研究を、分布に基づくアプローチ (分布クラスタリングや分布類似度 [11]) と語彙統語パターンの統合利用という観点から見ると、以下のような研究と関連が深い。Lin らは、分布類似度は類義語と反義語を区別できないという問題に着目して、分布類似度の高い単語を類義語と反義語に分類するために語彙統語パターンを使った [16]。Pantel らは、分布類似度をもとにクラスタリングされた単語集合に上位語を割り当てるために語彙統語パターンを用いた [21]。Mirkin らは、分布類似度と語彙統語パターンの情報を素性を使って、含意関係を認識する分類器を学習した [19]。

8 まとめ

本論文では，制約付きクラスタリングの枠組みを用いて，単語のクラスタリングに制約を導入した．また，語彙統語パターンを用いて制約を自動的に導出する手法も併せて提案した．実験の結果，制約を用いない場合に比べて，クラスタリングの質が向上することを確認した．

参考文献

- [1] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦 (編): 日本語語彙大系 (1997)
- [2] Baker, L. and McCallum A.: Distributional Clustering of Words for Text Classification, In Proceedings of SIGIR, (1998)
- [3] Basu S., Banerje A., Mooney R.: Semi-supervised Clustering by Seeding, In Proceedings of ICML, pp. 19-26, (2002)
- [4] Basu S., Bilenko M., and Mooney, J. R.: A Probabilistic Framework for Semi-Supervised Clustering, In Proceedings of SIGKDD, pp. 59-68, (2004)
- [5] Bilenko M., Basu S., and Mooney, J. R.: Integrating Constraints and Metric Learning in Semi-Supervised Clustering, In Proceedings of ICML, pp. 81-88, (2004)
- [6] Danushka B., Matsuo Y., and Ishizuka M.: An integrated Approach to Measuring Similarity between Words Using Information available on the Web, In Proceedings NAACL-HLT, pp. 340-347, (2007)
- [7] Brown, P. F., Pietra, D. J. V., deSouza, V. P., and Mercer, L. R.: Class-Based n-gram Models of Natural Language, Computational Linguistics, Vol. 18, No. 4, pp. 467-479, (1992)
- [8] Chklovski T. and Pantel P.: VERBOCEAN: Mining the Web for Fine-Grained Semantic Verb Relations, In Proceedings of EMNLP, pp. 33-40, (2004)
- [9] Fujita A., Inui K., and Matsumoto Y.: Detection of Incorrect Case Assignments in Automatically Generated Paraphrases of Japanese Sentences, In Proceedings of IJCNLP, pp. 14-21, (2004)
- [10] Hearst M.: Automatic Acquisition of Hyponyms from Large Text Corpora, In Proceedings of COLING, pp. 539-545, (1992)
- [11] Hindle D.: Noun Classification from Predicate-Argument Structures, In Proceedings of ACL, pp. 268-275, (1990)
- [12] Hofmann T.: Probabilistic Latent Semantic Indexing, In Proceedings of SIGIR, (1999)
- [13] Komachi M., Iida R., Inui K., and Matsumoto Y.: Learning Based Argument Structure Analysis of Event-nouns in Japanese, In Proceedings of PA-CLING, pp 120-128 (2007)
- [14] Kurohashi S. and Sakai Y.: Semantic Analysis of Japanese Noun Phrases : A New Approach to Dictionary-Based Understanding, In Proceedings of ACL, pp. 481-488, (1999)
- [15] Lange T., Law, M. H.C. L., Jain, A. K., and Buhmann, J.M.: Learning With Constrained and Unlabelled Data, In Proceedings of IEEE conference on computer vision and pattern recognition, (2005)
- [16] Lin D., Shaojun Z., Qin Li., and Zhou, M.: Identifying Synonyms among Distributionally Similar Words, In Proceedings of IJCAI, (2003)
- [17] Lu Z. and Leen, K. T.: Semi-supervised Learning with Penalized Probabilistic Clustering, In Proceedings of NIPS, (2004)
- [18] Matsuo Y., Sakaki T., Uchiyama K., and Ishizuka M.: Graph-based Word Clustering using a Web Search Engine, In Proceedings of EMNLP, (2006)
- [19] Mirkin S., Dagan I., and Geffet M.: Integrating Pattern-based and Distributional Similarity Methods for Lexical Entailment Acquisition, In Proceedings of COLING-ACL Poster Sessions, pp. 579-586 (2006)
- [20] Nelson B. and Cohen I.: Revisiting Probabilistic Models for Clustering with Pair-wise Constraints, In Proceedings of ICML, (2007)
- [21] Pantel P., and Ravichandran D.: Automatically Labeling Semantic Classes, In Proceedings of HLT-NAACL, (2004)
- [22] Pereira F., Tishby N., and Lee L.: Distributional Clustering of English Words, In Proceedings of ACL, pp. 183-190, (1993)
- [23] Resnik P.: Using Information Content to Evaluate Semantic Similarity in a Taxonomy, In Proceedings of IJCAI, pp. 448-453, (1995)
- [24] Rooth M., Riezler S., Prescher D., Carrol G., and Beil F.: Inducing a Semantically Annotated Lexicon via EM-Based Clustering, In Proceedings of ACL, pp. 104-111, (1999)
- [25] Shental N., Bar-Hillel A., Hertz T., and Weinshall D.: Computing Gaussian Mixture Models with EM using Equivalence Constraints, In Proceedings of NIPS, (2003)
- [26] Takamura T., Inui T., and Okumura M.: Extracting Semantic Orientations of Words using Spin Model, In Proceedings of ACL, (2005)
- [27] Torisawa K.: An Unsupervised Learning Method for Associative Relationships between Verb Phrases, In Proceedings of COLING, (2002)