

半教師あり語義曖昧性解消のためのグラフスパース化

小 嵩 耕 平^{†1} 小 町 守^{†1}
新 保 仁^{†1} 松 本 裕 治^{†1}

語義曖昧性解消の技術は、機械翻訳における訳語選択や情報検索における語義付きのインデキシングなどの様々な応用が期待されている。一方で、グラフに基づく分類問題に対する手法において、グラフの構築方法は重要であることが報告されており、b-マッチンググラフによる新しいグラフスパース化を用いた構築方法が提案されている。語義曖昧性解消タスクにおけるグラフを用いた既存の研究では、グラフスパース化について注目した研究はない。グラフスパース化の方法と語義の分類手法の組み合わせを比較実験することで、b-マッチンググラフによるグラフスパース化が語義曖昧性解消タスクにおいて有用であることを示す。また、グラフに対してクラスタ仮説を満たす程度を量る尺度を導入することで、b-マッチンググラフがk-近傍グラフと比較して、よりクラスタ仮説を満たしていることを示した。

Graph Sparsification for Semi-supervised Word Sense Disambiguation

KOHEI OZAKI,^{†1} MAMORU KOMACHI,^{†1}
MASASHI SHIMBO^{†1} and YUJI MATSUMOTO^{†1}

Word sense disambiguation (WSD) is one of the central problem in natural language processing. It has many applications such as document indexing for information retrieval and word selection in machine translation. Meanwhile, it is reported that the graph construction influences graph-based semi-supervised learning methods, and a new approach for graph sparsification called b-matching graph has been proposed. However, in previous graph-based WSD, no research has focused on graph sparsification. We compare graph sparsification methods in conjunction with inference methods and show that b-matching graph produces better prediction accuracy in WSD. In addition, We introduce a metric to measure how well the cluster assumption holds and show b-matching graph satisfies cluster assumption much more than k-nearest neighbors graph.

1. はじめに

文に現れる語に対して、周辺の文脈などを用いることにより語義の曖昧性を機械的に解消する語義曖昧性解消の技術は、機械翻訳における訳語選択や情報検索における語義付きのインデキシングなどの様々な応用が期待されている。語義曖昧性解消タスクに対して、いままでに様々なアプローチが提案されてきた。それらの研究には大きく分けて教師あり¹⁾、教師なし²⁾そして半教師あり^{3),4)}、の3つのアプローチに分けられる。

教師あり学習における語義曖昧性解消は、高い分類精度を達成しているが、そのためには十分な量の教師データが必要となる。語義曖昧性解消タスクは教師データを用意するために必要なアノテーションのコストが高いため、応用に必要となる分類精度を得られるだけの教師データが存在しない場合も多い。教師なし学習における語義曖昧性解消は、コストのかかる教師データを必要としない一方で、必ずしも所望する語義のクラスタが得られるとは限らないという問題がある。本稿では、タスクの性質から、十分に教師データが得られない場合について考える。半教師あり学習による語義曖昧性解消では、得ることができる少量の教師データのみを用いた教師あり語義曖昧性解消と比較して、アノテーションされていないラベルなしデータを加えて用いることで、より高い分類精度を得られることが期待できる。

半教師あり学習は近年盛んに研究されている分野である。中でも、データ間の類似関係により定義されるグラフに基づくアプローチは、他の半教師あり学習におけるアプローチと比較して、計算量の観点から大規模データに適用しやすいという利点があり、特に注目を集めている。自然言語処理における様々なタスクにおいても、グラフに基づくアプローチが有効であるという報告がなされている。^{4),5)}

一方で、グラフに基づく半教師あり学習においては、分類精度の向上と計算の効率化を目的として、グラフ構築のプロセスにおいて辺をカットする**グラフスパース化**がよく行なわれており、グラフスパース化の方法がパフォーマンスに大きな影響を与えることが報告されている。⁶⁾ グラフスパース化には一般的にk-近傍グラフがよく用いられるが、最近の研究ではb-マッチンググラフを用いることが提案されており、教師なしクラスタリングや半教師あり分類問題などにおいてその有効性が示されている。^{7),8)}

本稿では、半教師あり語義曖昧性解消タスクにおいてもb-マッチンググラフを用いたグ

^{†1} 奈良先端科学技術大学院大学 情報科学研究科

Graduate School of Information Science, Nara Institute of Science and Technology

ラフスパース化が有効であることを確認する。また、既存の研究では、分類問題におけるラベル伝播法 (LP) とその派生アルゴリズムにおいてのみ b-マッチンググラフによるグラフスパース化の効果が検証されてきたが、グラフを離散時間マルコフ連鎖 (MC) として用いるランダムウォークに基づくアプローチにおいても、b-マッチンググラフによるグラフスパース化の恩恵を得られることを確認し、それぞれの手法とグラフスパース化の組み合わせについて比較した。更に、グラフに対して、導出されるアルゴリズムの根拠となっているクラスタ仮説をどの程度満たすかを量る尺度の導入により、b-マッチンググラフが k-近傍グラフと比較して、よりクラスタ仮説を満たしていることを示した。

本稿の構成は以下の通りである。2 節にて関連研究をまとめ、3 節にて語義曖昧性解消タスクを半教師あり学習の文脈で定式化する。4 節にて半教師あり学習のアルゴリズムについて記述する。5 節では評価データセットを用いた実験結果を与え、グラフスパース化の比較をする。最後に 6 節にて我々の今回の仕事をまとめ、今後の課題について述べる。

2. 関連研究

半教師ありの語義曖昧性解消タスクにおいて、Bootstrapping やラベル伝播法 (LP) など様々なアプローチがいままで研究されてきた。^{3),4)} Niu らは、ラベルありデータがラベルなしデータと比較して少量しか得られない場合においては、Bootstrapping や Bilingual Bootstrapping などを用いた従来の半教師ありの手法^{3),9)}、あるいは SVM を用いた教師ありの手法¹⁰⁾ と比較して、グラフスパース化に k-近傍グラフを用いた LP による方法が、より優れた分類性能を発揮することを示した。⁴⁾

本研究ではこの結果を元に、LP などのグラフを用いた推論アルゴリズムにおいては、用いるグラフによってそのパフォーマンスが大きく左右されるために、グラフの構築が重要であることを示した最近の研究^{7),8)} に着目し、Niu らのアプローチからグラフスパース化の方法を変えることで、更なる分類精度の改善が見込めることを確認した。

Niu らは、k-近傍グラフを用いて LP による語義曖昧性解消を行ったが、我々は、グラフスパース化に着目し、グラフスパース化の方法と推論アルゴリズムの組み合わせについて比較した点が異なる。また b-マッチングに基づくグラフスパース化と、グラフを MC として扱う識別的ランダムウォーク (DRW)¹¹⁾ を自然言語処理分野において適用した事例はいままでになく、またその両者の組み合わせを検証したことも新しい試みである。

3. 半教師あり学習による問題の定式化

本節では、語義曖昧性解消タスクを半教師ありの分類問題に定式化する方法について述べる。語義に曖昧性のある語を w とし、この語が現われる文脈の集合を $X = \{x_i\}_{i=1}^n$ とする。 x_i は i 番目に w が現われる文脈であり、 n は文脈の総数である。 $S = \{s_j\}_{j=1}^c$ は曖昧性のある語 w における、割り当てが可能な語義ラベルの集合である。はじめの l 個のサンプル x_g ($1 \leq g \leq l$) をラベルありデータとし、他の u ($l+u = n$) 個のサンプル x_h ($l+1 \leq h \leq n$) をラベルなしデータとする。目的は、文脈 x_h に現れる語 w の語義ラベル ($\in S$) を、 x_g のラベル情報とサンプル集合 X の類似情報を使うことで推定することである。

X におけるクラスタ構造は、サンプルに対応する頂点を用いた連結グラフとして表すことができ、任意の 2 サンプル x_i, x_j 間の辺には、何らかの類似度尺度において頂点間がより似ているなら大きな重みを与えることにより得られる。ここで定義される頂点集合を V 、辺集合を E 、辺の重みを要素として持つ $n \times n$ 行列を W として定義することにより、重み付きグラフ $G = (V, E, W)$ が得られる。辺の重みについては、具体的には曖昧性のある語 w の文脈情報から特徴量を定め、サンプルを高次のベクトルとして表現したとき、グラフにおける辺の重みは、サンプル間のコサイン類似度などの類似度尺度を用いて決定する。

グラフを用いる半教師あり語義曖昧性解消では、この重み付きグラフ G を入力として、 x_h ($l+1 \leq h \leq n$) に対応するクラスタラベルを当てる問題として定式化される。

4. グラフに基づく半教師あり学習法

データ間の類似度によるグラフ G を用いて分類問題を解くアプローチは、グラフスパース化によるグラフの構築と、推論アルゴリズムの適用の 2 つのステップに分けることができる。これら 2 つのステップについてそれぞれ述べる。

4.1 グラフスパース化

重み付き完全グラフから、よりスパースな重み付き部分グラフを見つけるグラフスパース化は、重要なプロセスである。推論アルゴリズムが辺の数やスパース化された頂点の回数に従う計算量により定義されている場合は、グラフスパース化は計算の効率化に繋がる。また、データ間の類似度尺度によっては、近いデータにのみ類似度が有用であり、相対的に遠く離れたデータ間の類似度は信頼できない場合もあり、それらの類似度を無視することで、グラフスパース化による分類精度の向上も期待できる。グラフスパース化の手続きは、グラフ G の辺をカットすることにより行われる。任意の i, j 成分に対して、1 ならば $(i, j) \in E$ の辺

をカットせず、0 ならば $(i, j) \in E$ をカットすることを表現する行列として、 $n \times n$ の二値行列 P を導入し、グラフスパース化を行列 P を求める問題を考える。

k-近傍グラフはグラフスパース化の方法として広く使われてきた。^{4),5),12)} ある閾値 ε 以下の重みを持つ辺をカットする ε -近傍グラフが用いられることもあるが、この方法は孤立点を作りやすくパラメータの設定が困難な方法であるため、k-近傍グラフと比べると低いパフォーマンスを得る傾向にある。⁶⁾ そのため本稿においては、k-近傍グラフと b-マッチンググラフの2つのグラフスパース化について比較した。

k-近傍グラフは以下の最適化問題により定式化できる。

$$\max_{\hat{P}} \sum_{i,j} \hat{P}_{ij} W_{ij} \quad \text{s.t.} \quad \sum_j \hat{P}_{ij} = k, \hat{P}_{ii} = 0, \forall i, j \in 1, \dots, n \quad (1)$$

そして、すべての i, j に対して $P_{ij} = \max(\hat{P}_{ij}, \hat{P}_{ji})$ を求めることにより、二値行列 P を得る。この貪欲なアルゴリズムでは、得られるグラフの次数が $\sum_j P_{ij} \geq k$ と均一でないことから、クラスタ間を結ぶ辺が作られやすいと考えられる。

そこで、k-近傍グラフに代る手法として、より正則なグラフに近い b-マッチンググラフを分類問題に応用する提案がなされている。⁷⁾ b-マッチンググラフは以下の最適化を行なうことにより得られるグラフである。

$$\max_P \sum_{i,j} P_{ij} W_{ij} \quad \text{s.t.} \quad \sum_j P_{ij} = k, P_{ii} = 0, P_{ij} = P_{ji}, \forall i, j \in 1, \dots, n \quad (2)$$

この最適化により得られる行列 P は、対称性の制約を持つため、得られるグラフの次数が均一になる特徴がある。この最適化はループあり信念伝播法を用いることで効率的に近似解を求めることができる。^{7),13)}

以降の推論アルゴリズムにおける記述では、スパース化されたグラフに対して考える。すなわち任意の i, j に対して、 $W_{ij} \leftarrow W_{ij} P_{ij}$ とする。

4.2 推論アルゴリズム

本節では、グラフを用いた推論アルゴリズムの中で、重み付きグラフ G について定義されるコスト関数を最適化することによりクラスラベルを推定するアルゴリズムと、グラフをマルコフ連鎖 (MC) として扱い、MC 上でのランダムウォークに基づき未知のクラスラベルを推定するアルゴリズムについて、それぞれ述べる。

なお D は対角行列であり、各成分は $d_{ii} = \sum_j W_{ij}$ により定義されるものとする。またグラフ G において、ラベルありデータに関連付けられている頂点を V_l とし、ラベルなしデータに関連付けられている頂点を V_u とする。

4.2.1 ラベル伝播法 (LP)

LP は、以下の式に定義されるコスト関数を最小化することにより、最適な識別関数 F を求める手法である。¹⁴⁾

$$\min_{F \in R^{|V| \times |S|}} \text{tr}\{F^T \Delta F\} \quad (3)$$

ただし、 Δ はグラフラプリアン $\Delta = D - W$ である。 F は $|V| \times |S|$ の実数値行列であり、各成分は頂点 $v_i \in V$ とクラスラベル $s_j \in S$ との間隔を表す量を持つ。また、 F をラベルあり頂点 V_l とラベルなし頂点 V_u に対応する行により分割した行列をそれぞれ F_l, F_u とする。いまラベルありデータの頂点とそのクラスラベルに対応する成分を 1 とし、それ以外を 0 として定義される $|V_l| \times |S|$ 行列を Y_l とする。上式における最適化はラベルなし頂点に $\Delta F_u = 0$ の、ラベルあり頂点に $F_l = Y_l$ の2つの制約の元で行なう。

この最適化により得られた F_u において、各頂点 $v_i \in V_u$ と間隔を最大とするクラスラベル $s_j \in S$ がラベルなし頂点のクラスラベルとして推定される。

4.2.2 識別的ランダムウォーク (DRW)

グラフにおけるランダムウォークは、訪問する頂点の列により記述される MC によってモデル化される。MC における各状態はグラフの頂点に対応する。MC の遷移確率行列は、 $P = D^{-1}W$ により与えられる。

DRW による手法では、長さの上限が定められたランダムウォークに基づく尺度を導入し、この尺度において、頂点と間隔を最大とするようなクラスラベルを頂点に割り当てることで、ラベルなし頂点の分類を行なう。¹¹⁾ LP がラベルありデータ間の辺をすべて無視してクラスラベルを推定する性質を持つことに対して、DRW ではより多くの辺を考慮した推定を行なうため、周辺のラベルありデータの分布をより良く反映する。このためクラスタ仮説がよく満たされている場合は、より良い分類精度が得られると期待できる。

定義 4.1 (D-walk) 与えられた MC に対して、あるクラスラベル $y \in S$ と、離散的な長さ $l > 1$ を考える。このとき D-walk は状態の列 q_0, \dots, q_l として定義される。ただし、すべての $0 < t < l$ に対して、 $y_{q_0} = y_{q_l} = y$ かつ $y_{q_t} \neq y$ を満たす。

ここで、長さ l のすべての D-walk の集合を D_l^y と表記する。また、与えられた L 以下の長さにより定義される。D-walk の集合を $D_{\leq L}^y$ と表記する。

定義 4.2 (上限付き D-walks betweenness) 与えられたラベルなし頂点 $q \in V_u$ とクラスラベル $y \in S$ における上限付き D-walks betweenness 関数は、 q と y の間隔を表す量として、次のように定義される。

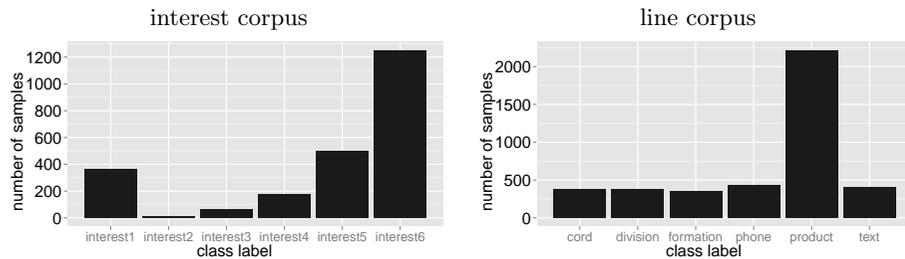


図1 クラスラベルの分布

$$B_L(q, y) = E[\text{pt}(q) | D_{\leq L}^y] \quad (4)$$

ただし $\text{pt}(q)$ は頂点 q が訪問された回数を数える関数である。

DRW では考慮するウォークの長さに上限を加えることで、上限のない場合と比較して良い分類精度を得ることができると報告されている。また $B_L(q, y)$ は、隠れマルコフモデルのパラメータ推定に用いられる Baum-Welch アルゴリズムに類似した前向き後ろ向きの再帰計算を用いることで、時間計算量を $\Theta(|E|L)$ とする、効率的な計算が可能となる。最終的に、ラベルが未知の頂点 q に対して、 $B_L(q, y)$ を最大にするクラスラベル $y \in S$ を割り当てることにより、クラスラベルの推定が行われる。

5. 語義曖昧性解消タスクにおけるグラフスパース化の評価実験

5.1 評価データセット

語義曖昧性解消タスクにおけるグラフスパース化の比較のために、我々は英語の名詞における語義曖昧性解消タスクに広く用いられている評価データセット*1 を用いて実験を行った。このデータセットは、主に Wall Street Journal の記事から “interest,” “line” の語が用いられる文脈を取り出し、タグ付けをしたものである。¹⁵⁾ “interest,” “line” コーパスには、それぞれ 3103, 4141 のサンプルが用意されている。教師データが少量しか得られない状況を人工的に作るために、1割のデータをラベルありデータとして、残りをラベルなしデータとして扱う。それぞれの語は複数のクラスラベルを持ち、各クラスラベルの出現頻度には大きな偏りがある(図1)。語義ラベルの出現頻度の偏りはよく現れるので、最頻出語義を常に分類結果として答える分類器がベースラインとしてしばしば設けられる。

*1 <http://www.d.umn.edu/~tpederse/data.html>

表1 各スパース化の方法と手法の組み合わせによる英語語義曖昧性解消タスクのパフォーマンス比較

Ambiguous words	with LP			
	BM	KNN	FULL	MFS (baseline)
interest	83.4%	81.9%	56.7%	53.0%
line	67.8%	66.6%	53.6%	53.6%

Ambiguous words	with DRW			
	BM	KNN	FULL	MFS (baseline)
interest	83.3%	81.9%	61.4%	53.0%
line	72.1%	70.0%	53.6%	53.6%

5.2 実験結果

比較のため、特徴量の取り方と類似度尺度を Niu らによる先行研究⁴⁾と同様のものを用いた。特徴量は、曖昧性を解消したい語の周辺の語の品詞とその語の位置、順序を考慮しない周辺文脈に出現する語、そして局所的なコロケーションを用いた。¹⁶⁾ また類似度はコサイン尺度を用いた。各スパース化により得られたグラフを用いて LP と DRW による語義曖昧性解消を評価した。

表1は各推論アルゴリズムとグラフスパース化の手法に対して、語義曖昧性解消タスクにおける分類精度を計算した結果である。表における BM, KNN はそれぞれ、グラフスパース化に b-マッチンググラフ, k-近傍グラフを用いたことを示している。FULL はグラフスパース化を行わず、完全グラフの状態 で推論アルゴリズムを適用し分類したことを示す。MFS (Most Frequent Sense) は最頻出語義を分類結果として常に返すベースラインによる分類を示す。DRW におけるパラメータは $L = 5$ と固定した。k-近傍グラフにおける k の値, b-マッチンググラフにおける b の値は 3 から 40 までの範囲に定め、得られたすべてのグラフに対して推論アルゴリズムを適用し、得られた精度の最も高い結果を用いた。パラメータ k, b を変動させたときの分類精度の変化をグラフにしたものが図2である。上段のグラフは推論アルゴリズムとして LP を用いた結果であり、下段のグラフは DRW を用いた結果である。ほとんどの場合において b-マッチンググラフによるスパース化を行った場合の分類精度が k-近傍グラフによる場合と比べて高い分類精度を達成していることが確認できる。

5.3 クラスタ仮説に基づくグラフスパース化の比較

実験に用いた LP や DRW はクラスタ仮説に基づいて設計されている。クラスタ仮説とは、似たサンプルは同じクラスラベルに割り当てられやすいという仮説である。言い換えると、LP や DRW においてはグラフがクラスタ仮説を満たしていなければ、高い分類精度を

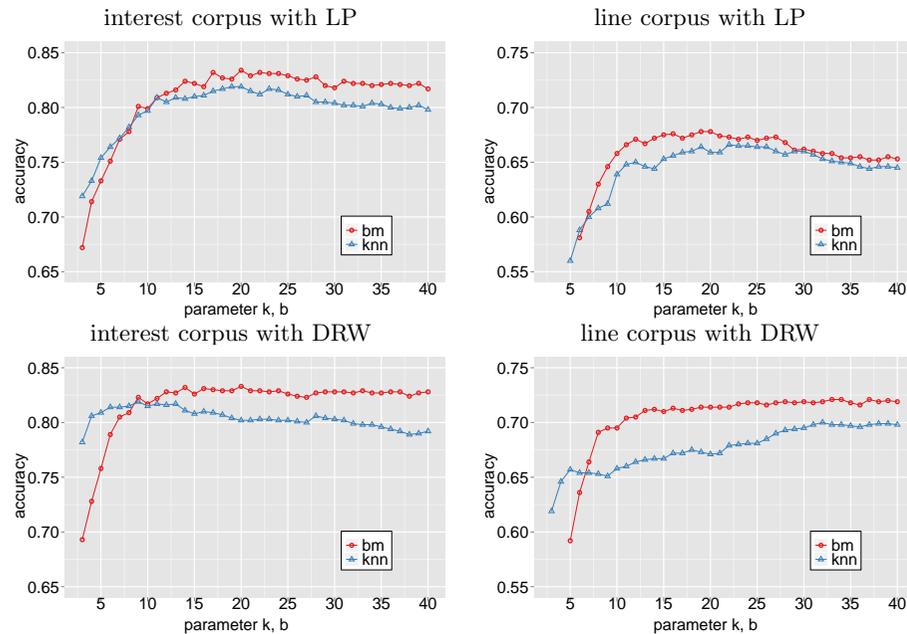


図 2 分類精度とスパース化のサイズを軸に各グラフスパース化の方法を比較

期待することができない。そこで、各グラフスパース化のどちらがこの仮説を満たしているかを量ることで、グラフを比較する。

$l(i)$ を頂点 i に対するクラスラベルとする。以下のような指標を定義する。この量を **グラフ G に対するクロスラベルエラー** とする。

$$E_{cle}(G) = \frac{\sum_{(i,j) \in E} W_{ij} 1[l(i) \neq l(j)]}{\sum_{(i,j) \in E} W_{ij}} \quad (5)$$

ただし、 $1[l(i) \neq l(j)]$ は $l(i) \neq l(j)$ のときのみ 1 をとり、 $l(i) = l(j)$ のときは 0 をとる指示関数である。この量はグラフ中のすべての辺において、異なるクラスラベルをむすぶ辺の重みの割合に相当する。この指標が最大値/最小値をとる場合の例を図 3 に示す。

いま、グラフのクラスラベルがすべて既知であるとして、各スパース化の方法によるグラフに対してクロスラベルエラーを比較した結果が図 4 である。図における横軸はパラメー

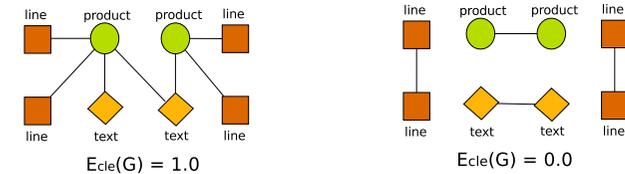


図 3 グラフに対するクロスラベルエラーが最大値と最小値をとる例

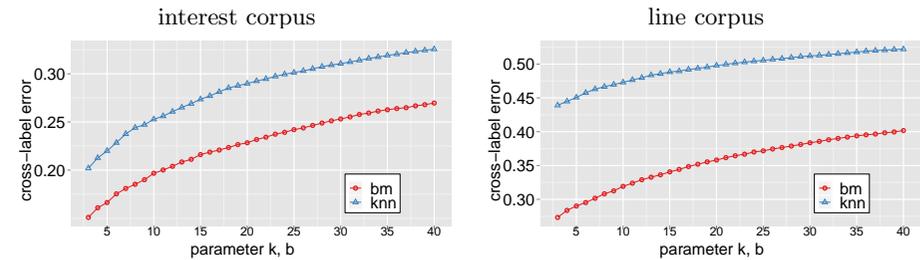


図 4 グラフに対するクロスラベルエラー

タ k, b の値を用いた。この結果から b -マッチンググラフにおいてグラフスパース化を行った場合、 k -近傍グラフと比較してより小さなクロスラベルエラーであることがわかる。

クロスラベルエラーは大きいほど誤分類を起こしやすい。このことはグラフに対するクロスラベルエラーの指標を頂点に対して定義し、誤分類を起こす頂点に対するクロスラベルエラーと、正しく分類される頂点に対するクロスラベルエラーがどのように分布するかを見ることにより確認する。以下のように **頂点 v に対するクロスラベルエラー** を定義する。

$$E_{cle}(v) = \frac{\sum_{(v,t) \in E} W_{vt} 1[l(v) \neq l(t)]}{\sum_{(v,t) \in E} W_{vt}} \quad (6)$$

図 5 は “interest” データセットにおいて、 k -近傍グラフと b -マッチンググラフを LP に用いたときの正例と負例それぞれに対応する頂点におけるクロスラベルエラーがどのように分布しているかを示す図である。パラメータ k, b は共に LP を用いたときに分類精度が最も良かった $k = b = 20$ における結果を用いた。データの分布を確認するために、図の横軸には頂点の次数を用いた。また、 b -マッチンググラフでは頂点の区別のために、図の水平方向に標準正規分布による小さなノイズを加えた。この図からは、クロスラベルエラーが高

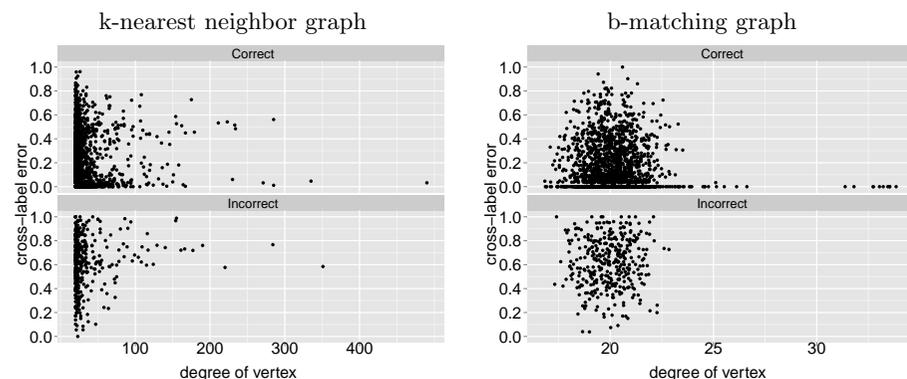


図5 異なる頂点度数における頂点に対するクロスラベルエラーの分布

いほど誤分類が起こる傾向にあることが確認できる。以上の結果から、b-マッチンググラフを用いることで、k-近傍グラフを用いる場合と比較して、クラスタ仮説をより満たすグラフが得られるため、より高い分類精度が得られたことが見て取れる。

6. まとめ

語義曖昧性解消タスクにおいて、グラフスパース化にb-マッチンググラフを用いた場合には、従来用いられてきたk-近傍グラフと比較して分類性能が上回ることを確認した。また、いままで自然言語処理において応用されていなかったDRWに対してもb-マッチンググラフによりグラフスパース化をすることで、k-近傍グラフにおける結果と比較して分類性能が上回ることを確認した。b-マッチンググラフとDRWの組み合わせにより、データセットによっては分類精度を大きく改善できることが実験結果から見て取れる。

また、本稿ではクラスタ仮説から導出されるアルゴリズムに適したグラフであるかを測る指標として、グラフに対するクロスラベルエラーを定義した。クロスラベルエラーを用いてb-マッチンググラフにより構築されるグラフがk-近傍グラフと比較して、より良くクラスタ仮説を満たしていることを確認した。b-マッチンググラフを用いることで、k-近傍グラフと比較してクラスタ仮説をより満たすグラフが得られる理論的根拠、あるいはよりクラスタ仮説をよく満たすグラフにスパース化する手法の導出が今後の課題である。

参考文献

- 1) Leacock, C., Miller, G.A. and Chodorow, M.: Using corpus statistics and WordNet relations for sense identification, *Computational Linguistics*, Vol.24, No.1, pp. 147–165 (1998).
- 2) Schütze, H.: Automatic word sense discrimination, *Computational Linguistics*, Vol.24, No.1, pp.97–123 (1998).
- 3) Yarowsky, D.: Unsupervised word sense disambiguation rivaling supervised methods, *Proc. of ACL-1995*, pp.189–196 (1995).
- 4) Niu, Z.-Y., Ji, D.-H. and Tan, C.L.: Word sense disambiguation using label propagation based semi-supervised learning, *Proc. of ACL-2005*, pp.395–402 (2005).
- 5) Goldberg, A.B. and Zhu, X.: Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization, *Proc. of TextGraphs: the First Workshop on Graph Based Methods for Natural Language Processing*, pp. 45–52 (2006).
- 6) Maier, M., Luxburg, U.V. and Hein, M.: Influence of graph construction on graph-based clustering measures, *Proc. of NIPS-09*, pp.1025–1032 (2009).
- 7) Jebara, T., Wang, J. and Chang, S.-F.: Graph construction and b-matching for semi-supervised learning, *Proc. of ICML-2009*, pp.441–448 (2009).
- 8) Jebara, T. and Shchogolev, V.: B-Matching for spectral clustering, *Proc. of ECML*, pp.679–686 (2006).
- 9) Li, H. and Li, C.: Word translation disambiguation using bilingual bootstrapping, *Computational Linguistics*, Vol.30, No.1, pp.1–22 (2004).
- 10) Mihalcea, R., Chklovski, T. and Kilgarriff, A.: The Senseval-3 English lexical sample task, *Proc. of Senseval-3: the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pp.25–28 (2004).
- 11) Callut, J., François, K., Saerens, M. and Dupont, P.: Semi-supervised classification from discriminative random walks, *Proc. of ECML/PKDD*, pp.162–177 (2008).
- 12) Szummer, M. and Jaakkola, T.: Partially labeled classification with markov random walks, *Proc. of NIPS-02*, pp.945–952 (2002).
- 13) Huang, B.: Loopy belief propagation for bipartite maximum weight b-matching, *Proc. of AISTATS*, pp.195–202 (2007).
- 14) Zhu, X., Ghahramani, Z. and Lafferty, J.D.: Semi-supervised learning using gaussian fields and harmonic functions, *Proc. of ICML-2003*, pp.912–919 (2003).
- 15) Pedersen, T.: A simple approach to building ensembles of Naive Bayesian classifiers for word sense disambiguation, *Proc. of NAACL-2000*, pp.63–69 (2000).
- 16) Lee, Y.K. and Ng, H.T.: An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation, *Proc. of EMNLP-2002*, pp.41–48 (2002).