

## 素性間の共起性を検査する Co-training による語義判別規則の学習

新納浩幸

茨城大学 工学部 システム工学科  
shinnou@dse.ibaraki.ac.jp

自然言語処理では個々の問題を分類問題として定式化し、帰納学習の手法を利用して、その問題を解決するというアプローチが大きな成功をおさめている。しかしこのアプローチには帰納学習で必要とされる訓練データを用意しなければならないという大きな問題がある。この問題に対して、近年、少量のラベル付き訓練データから得られる分類規則の精度を、大量のラベルなし訓練データによって高めてゆく seed 型の学習が散見される。ここではその中心的な手法である Co-training を語義判別規則に適用することを試みる。ただし Co-training では独立な 2 組の素性集合を設定する必要がある。現実的にはこの独立性の条件が厳しいため、得られる規則の精度が頭打ちになってゆく。本論文ではこの問題を回避するために、追加事例の選択に素性間の共起性を考慮することで Co-training の手法を改良する。実験では 3 つの語義選択問題について本手法を適用した。結果、通常の Co-training を適用する以上の精度の向上が見られた。

## Learning of word sense disambiguation rules by Co-training checking co-occurrence of features

Hiroyuki Shinnou

Department of Systems Engineering, Ibaraki University  
shinnou@dse.ibaraki.ac.jp

In natural language processing, it is effective to convert problems to classification problems, and to solve them by an inductive learning method. However, this strategy needs labeled training data which is fairly expensive to obtain. To overcome this problem, some learning methods using unlabeled training data have been proposed. Co-training is representative in such methods. In this paper, we apply the co-training method to a word sense disambiguation problem. However, co-training needs two features which satisfy the conditional independence assumption. This assumption is too rigid. We have no choice but to use incomplete two features, and then the accuracy of learned rules reaches the ceiling. In this paper, we check co-occurrence between two features to avoid such phenomena when we add unlabeled instances to training data. We applied our method to word sense disambiguation problems for three words. Our method improved the precision of the original co-training.

# 1 はじめに

本論文では、追加事例の選択に素性間の共起性を考慮することで Co-training の手法を改良する。Co-training は少量のラベル付き訓練データと大量のラベルなし訓練データから分類問題を解決する手法である。本手法を語義選択問題に応用し、少量のラベル付き訓練データから語義選択規則を学習する。

自然言語処理では個々の問題を分類問題として定式化し、帰納学習の手法を利用して、その問題を解決するというアプローチが大きな成功をおさめている。しかしこのアプローチには帰納学習で必要とされるラベル付き訓練データを用意しなければならないという大きな問題がある。一般に、ラベル付き訓練データが大規模であるほど精度の良い規則を学習できるが、データに対するラベル付けは手作業で行わなくてはならないため、ラベル付き訓練データの作成のコストが高い。

この問題に対して、近年、少量のラベル付き訓練データから得られる分類規則の精度を、大量のラベルなし訓練データによって高めてゆく seed 型の学習が試みられている [11, 7, 6, 8]。その中でも特に Co-training という学習手法 [2] は、ラベルなし訓練データを用いることで、分類規則の精度向上がなされることを、PAC 学習の枠組みで理論的に示している点で注目されている。Co-training は独立な 2 つの素性集合を設定し、一方の素性集合のみを用いてラベル付き訓練データから分類器 1 を作成する。分類器 1 を用いてラベルなし訓練データの判別を行い、信頼性の高いものをラベル付き訓練データに加える。同様にして、もう一方の素性集合のみを用いることによって、ラベルなし訓練データの一部をラベル付き訓練データに加える。このようにラベル付き訓練データを増やすことで分類器の精度を向上させてゆく。ここで分類器の精度が向上してゆく仕掛けは、2 つの素性集合が独立であるという条件である。この条件は、一方の素性から判断して追加されてゆく事例が、もう一方の素性集合からみるとランダムであることを保証しているために、精度が向上する。Co-training は独立な 2 つの素性集合さえ設定できれば、実装は容易であるために、文書分類 [2]、固有表現抽出 [4] に応用されている。

ただし Co-training が仮定する独立性をもつ素性集合を設定することは難しい。現実的には、ある程度、この条件を満たすしかない。そのような素性集

合でも、Co-training を利用することで、分類規則の精度は向上する。しかし学習できる規則の精度は頭打ちになってゆく。本論文ではこの原因について概観し、その打開策として追加する事例の選択において、素性間の共起性を考慮することを提案する。これによって通常の Co-training で得られる規則の精度を更に高めることができる。

実験では 3 つの単語を対象に、語義判別規則の学習を、Co-training により行った。通常の Co-training に提案手法を組み合わせることで、分類精度を大きく向上させることができた。

## 2 Co-training

Co-training のアルゴリズムを以下に示す [2]。

step 0 少量のラベル付き訓練データ  $L$  と大量のラベルなし訓練データ  $U$  を用意する。

step 1  $U$  から  $u$  個の事例をランダムに取り出し  $U'$  を作る。

step 2 素性集合  $x_1$  だけに注目して  $L$  から分類器  $h_1$  を学習する。

step 3 素性集合  $x_2$  だけに注目して  $L$  から分類器  $h_2$  を学習する。

step 4  $h_1$  を使って  $U'$  の事例にラベルをつけ、信頼性の高いものから  $p$  個の正例と  $n$  個の負例を選ぶ。

step 5  $h_2$  を使って  $U'$  の事例にラベルをつけ、信頼性の高いものから  $p$  個の正例と  $n$  個の負例を選ぶ。

step 6 step 4 と 5 で得られたラベル付きの事例  $2p + 2n$  個を  $L$  に加える。

step 7 step 1 から繰り返す。

上記アルゴリズムの  $u, p, n$  の数は問題に応じて決める。

Co-training が有効に機能する要は step 4 と step 5 である。step 4 において  $h_1$  を使って取り出される事例が  $x_2$  からみればランダムであるために、それら事例を追加した  $L$  と素性集合  $x_2$  から  $h_2$  を作れば、その精度は前回の  $h_2$  の精度よりも良く

なる．step 5 も同様に， $h_1$  の精度を向上させる．このように互いにもう片方の素性に有益な事例を追加し合うことで Co-training が有効に機能する．

Co-training で  $x_1$  と  $x_2$  に仮定されている素性集合の独立性は，step 4 と step 5 において取り出される事例がもう片方の素性集合にとって，ランダムであることを保証している．このことの概略を示す．

Co-training では，事例の集合  $X$  は 2 次元とみなせる．つまり  $X = (X_1, X_2)$  と表せる． $D$  を  $X$  上の分布， $X$  上の目的関数を  $f$ ， $X_1$  上の目的関数を  $f_1$ ， $X_2$  上の目的関数を  $f_2$  とする．

素性集合  $x_1$  と  $x_2$  が独立であることの定義は， $f_1$ ， $f_2$ ， $D$  が，任意の  $(\hat{x}_1, \hat{x}_2) \in X$  に対して，以下の 2 つの条件式を満たすことである．

$$\begin{aligned} Pr[x_2 = \hat{x}_2 | x_1 = \hat{x}_1] \\ = Pr[x_2 = \hat{x}_2 | f_1(x_1) = f_1(\hat{x}_1)] \quad \dots (1) \end{aligned}$$

$$\begin{aligned} Pr[x_1 = \hat{x}_1 | x_2 = \hat{x}_2] \\ = Pr[x_1 = \hat{x}_1 | f_2(x_2) = f_2(\hat{x}_2)] \quad \dots (2) \end{aligned}$$

step 4 と step 5 は  $x_1$  と  $x_2$  に関して対称であるので，以後の説明は step 4 に関してだけ行う．

step 4 において得られたある正の事例を  $(\hat{x}_1, \hat{x}_2)$  とする．式 1 の左辺は，以下の集合  $A$  から，適当に事例を取り出したときにそれが  $(\hat{x}_1, \hat{x}_2)$  である確率を表している．

$$A = \{(x_1, x_2) \in D | x_1 = \hat{x}_1\}$$

また式 1 の右辺は， $D$  の中の正の事例から，適当に事例を取り出したときに，それが  $(\hat{x}_1, \hat{x}_2)$  である確率を表している．つまり事例  $(\hat{x}_1, \hat{x}_2)$  は， $x_2$  から見れば，正の事例の中からランダムに取り出した事例とみなせる．同様に，step 4 においては， $x_2$  から見れば負の事例の中からランダムに与えられた事例も取り出される．追加される正の事例の数  $p$  と負の事例の数  $n$  を， $D$  の中の正の事例の数と負の事例の数の割合から設定しておけば，追加される  $p+n$  個の事例は， $x_2$  から見た場合，平均的に  $D$  の中からランダムに取り出されたものと見なすことができる．

### 3 素性間の共起性を利用した事例の追加

現実には，独立性を満たすような 2 組の素性集合を設定することは難しい．特に自然言語処理では，

素性どうしには共起性が存在することが多く，独立性の条件を破ってしまう． $\hat{x}_1$  と  $\hat{x}_2$  の共起性が強いとき，

$$Pr[x_2 = \hat{x}_2 | x_1 = \hat{x}_1]$$

の値は他の  $(x_2, \hat{x}_1)$  を取り出す確率と比べ高くなり，式 1 の条件を満たさなくなることは明らかである．実際の問題では，step 4 において  $(\hat{x}_1, \hat{x}_2)$  が  $L$  に追加された場合，その追加された  $L$  から学習される次の  $h_2$  では， $\hat{x}_2$  の判定の信頼性が高くなる．この結果， $\hat{x}_2$  と共起しやすい  $\hat{x}_1$  を  $X_1$  の要素としてもつ事例  $(\hat{x}_1, \hat{x}_2)$  が再び  $L$  に追加される．これが繰り返されると， $x_1$  と  $x_2$  のどちらにとっても，新たな事例の追加がない状況に陥り， $h_1$  や  $h_2$  の精度が頭打ちになる．

このような状況を打開するために，本論文では事例  $(\hat{x}_1, \hat{x}_2)$  を追加する際に， $L$  の分布から  $\hat{x}_1$  と  $\hat{x}_2$  の共起性を調べ，共起性がある程度高い場合には，これ以上その事例を加えてもその後の  $h_1$  や  $h_2$  の学習に効果がないと判断し，その事例は追加せずに，次候補の事例を追加する戦略をとる．

注意として，素性  $\hat{x}_1$  と素性  $\hat{x}_2$  との共起性は，事例全体の集合  $D$  における共起性ではなく，ラベル付きの訓練データ  $L$  における共起性である．また素性集合間の共起性の測定方法は，設定する素性集合に応じて適切なものを定義する必要がある．一般に共起性の測定方法として，相互情報量や dice 係数など幾つかの方法が提案されているので，それらを拡張して利用できる．

## 4 語義判別問題への応用

本論文の提案手法を語義選択問題に応用する．語義選択問題はある文脈  $b$  上に現れた単語  $w$  の語義が正か負かを判定する問題である<sup>1</sup>．これは分類問題そのものである．文脈  $b$  からどのような素性を設定するかが解決のポイントである．

### 4.1 素性集合の設定

Co-training を利用するためには，ある程度独立と考えられる 2 組の素性集合が必要になる．この設定のために，単語  $w$  が現れた文脈  $b$  を左文脈  $b_l$  と右文脈  $b_r$  に分割する． $b_l$  とは  $w$  の左に位置する単

<sup>1</sup>ここでは  $w$  の語義は正と負の 2 種類であると仮定している．

表 1: 設定した素性

素性名	素性値
l1	1つ前の単語
l2	(2つ前の単語)-(1つ前の単語)の2単語列
l3	(3つ前の単語)-(2つ前の単語)-(1つ前の単語)の3単語列
r1	1つ後の単語
r2	(1つ後の単語)-(2つ後の単語)の2単語列
r3	(1つ後の単語)-(2つ後の単語)-(3つ後の単語)の3単語列

語列であり,  $b_r$  は  $w$  の右に位置する単語列である.  $b_l$  から得られる素性集合を  $x_1$  とし,  $b_r$  から得られる素性集合を  $x_2$  とする.

例えば「声」の語義は『意見』という語義と『喉から発声される音』という語義がある. 「日本国民の声を集めました」という文の「声」の左文脈は「日本国民の」であり, 右文脈は「を集めました」である.

次に  $x_1$  では, 表.1 に示す l1, l2, l3 の3つの素性を設定する. 同様に  $x_2$  では, 表.1 に示す r1, r2, r3 の3つの素性を設定する.

例えば「日本国民の声を集めました」という文は以下のように形態素解析される. 各行が分割された単語であり, 第1列が表記, 第2列が原型, 第3列が品詞を表す.

日本	日本	名詞-固有名詞-地域-国
国民	国民	名詞-一般
の	の	助詞-連体化
声	声	名詞-一般
を	を	助詞-格助詞-一般
集め	集める	動詞-自立
まし	ます	助動詞
た	た	助動詞

この結果から以下の6つの素性ができる.

l1 = の, l2 = 国民-の, l3 = 日本-国民-の,  
r1 = を, r2 = を-集める, r3 = を-集める-ます

このようにして設定した  $x_1$  と  $x_2$  は, ある程度独立な2組の素性集合になる.

ラベル付き訓練データ  $L$  を利用して, 素性集合  $x_1$  のみから, あるいは素性集合  $x_2$  のみから語義判別の規則を学習する必要がある. これらの学習に, 本論文では決定リスト [10] を利用する.

## 4.2 共起性の測定

ここでは素性集合  $\hat{x}_1$  と素性集合  $\hat{x}_2$  との共起性  $Cor(\hat{x}_1, \hat{x}_2)$  を以下のように定義する.

$$Cor(\hat{x}_1, \hat{x}_2) = \max_{i,j} dice(\hat{l}_i, \hat{r}_j)$$

ただし  $dice(\hat{l}_i, \hat{r}_j)$  は  $\hat{l}_i$  と  $\hat{r}_j$  の dice 係数であり, 以下で定義される.

$$dice(\hat{l}_i, \hat{r}_j) = \frac{2frq(\hat{l}_i, \hat{r}_j)}{frq(\hat{l}_i) + frq(\hat{r}_j)}$$

ここで  $frq(\hat{l}_i), frq(\hat{r}_j)$  はそれぞれ,  $L$  の中で現れた  $\hat{l}_i$  と  $\hat{r}_j$  の頻度であり,  $frq(\hat{l}_i, \hat{r}_j)$  は  $\hat{l}_i$  と  $\hat{r}_j$  が  $L$  の中で共起した頻度である.

例えば,  $L$  中の事例として, 以下の6つの素性があったとする.

l1 = の, l2 = 国民-の, l3 = 日本-国民-の,  
r1 = を, r2 = を-集める, r3 = を-集める-ます

このとき以下の頻度関係が得られる.

frq(l1 = の) = 1  
frq(l2 = 国民-の) = 1  
frq(l3 = 日本-国民-の) = 1  
frq(r1 = を) = 1  
frq(r2 = を-集める) = 1  
frq(r3 = を-集める-ます) = 1  
frq((l1 = の, r1 = を) = 1  
frq((l1 = の, r2 = を-集める) = 1  
frq((l1 = の, r3 = を-集める-ます) = 1  
...  
frq((l3 = 日本-国民-の,  
r3 = を-集める-ます) = 1

## 5 実験

単語「声」の語義判別規則の学習に本手法を適用する。「声」の語義は大きく『意見』と『喉から発声される音』という語義がある。ここでは曖昧性をなくするために『意見』という語義とそれ以外の語義という形で2つの語義を設定し、前者を正、後者を負とする。

次に毎日新聞の'93から'97年の5年分の記事を形態素解析し、「声」という単語を含む文を取り出した。全部で30,458文存在した。次に、そこからランダムに100文と500文を取り出し、それらの各文の持つ「声」の語義に応じて正または負のラベルを付与した。ラベル付きの100文を最初のラベル付き訓練データ  $L$  とし、ラベル付きの500文を評価のためのテストデータ  $T$  とした。残りの29,858文がラベルなし訓練データ  $U$  である。Co-trainingにおける各繰り返しで取り出す  $u$  の数は50に、また正例の数  $p$  と負例の数  $n$  はともに3とした。

また本論文で提案する素性集合間の共起性を測るには  $L$  から素性の頻度表を作成する必要があるが、 $L$  は Co-training の各繰り返しの中で徐々に増加してゆく。各繰り返しの時点で  $L$  から素性の頻度表を更新することも考えられるが、ここでは効率の問題から50回の繰り返し毎にこの頻度表を更新することにした。

追加する事例は基本的に信頼度の高いものが優先である。本論文では決定リストを用いているので、決定リストの順位を信頼度として利用できる。ただし追加する事例  $(\hat{x}_1, \hat{x}_2)$  の共起性  $Cor(\hat{x}_1, \hat{x}_2)$  の値が0.3以上である場合、この事例は追加せずに次に信頼度の高い事例を追加候補にする。

実験の結果を図1に示す。図の縦軸はテストセットによる正解率を表す。横軸は Co-training における繰り返しの回数を表す。曲線 original は通常の Co-training による正解率の推移を表し、曲線 our\_method は本手法を用いた場合の正解率の推移を表す。Co-training を用いることで、最初の100事例から学習できた規則の精度が大きく向上していることがわかる。また通常の Co-training では正解率0.77辺りで精度が頭打ちになっているが、本手法では更に精度が上がっていることも確認できる。

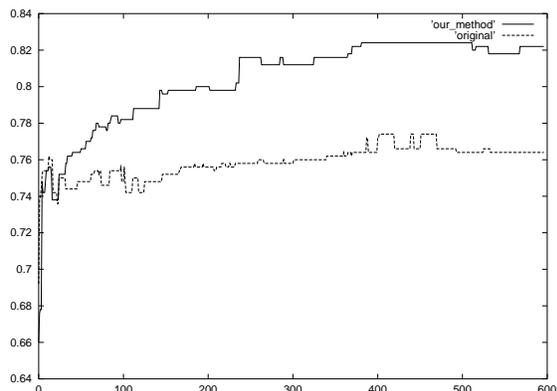


図1: 正解率の推移(「声」)

同様の語義判別の実験を「トップ」と「壁」について行った。「トップ」には大きく『先頭, 第1位』と『最高責任者』という語義があるが、ここでは『先頭, 第1位』の語義を正、それ以外の語義を負とした。また「壁」にも大きく『部屋と部屋との隔て』と『障害物』という語義があるが、ここでは『部屋と部屋との隔て』の語義を正、それ以外の語義を負とした。それぞれのデータ数や設定したパラメータを表2に示す。

表2: パラメータ

	声	トップ	壁
$ L $	100	100	100
$ U $	29,858	14,569	5,866
$ T $	500	300	100
$u$	50	50	50
$p$	3	3	3
$n$	3	1	2

「トップ」と「壁」についての実験結果を図2と図3に示す。どちらも「声」の場合と同様、通常の Co-training による学習よりも、より精度の高い規則を学習できた。

「声」「トップ」「壁」の各々実験で最も正解率の高かった値を取ったものが表3である。ベースラインは、最初の100事例の  $L$  から学習された決定リストの正解率である。

## 6 考察

ある分類問題に対して Co-training が利用できるかどうかのポイントは、ある程度独立な2つの素性集合を見つけられるかどうかである。ある程度独立な2つの素性集合は、実際はその問題を解決する手がかりとなる2つの独立した観点に対応する。WEB上の文書の文書分類問題に対しては、その文書内の単語とその文書をリンクしている文書という2つの観点がそれにあたる [2]。固有表現抽出問題に対しては、表現内の文字列とその表現があらわれる文脈がそれにあたる [4]。本論文では語義選択問題を扱い、左文脈と右文脈という2つの観点を導入した。これはある単語の語義は、その単語を修飾している単語列だけから予想でき、それとは別に、その単語を格としてもつ動詞から予想できるという考えが基本である。例えば、「日本国民の声」と言えば、その「声」の語義は『意見』であるとわかるし、「声を集めました」と言えば、その「声」の語義は『意見』であるとわかる。そのため本来は名詞が対象であるが、動詞の場合、連体修飾の用法も多いので、動詞についても適用できる可能性がある。しかし、本来、語義選択問題は対象とする単語ごとに、語義の判別に有効な素性が異なる。そのため、本論文で導入した左文脈と右文脈という2つの観点が有効な単語も限定されている。根本的には、語義選択問題に有効そうな多数の素性を準備し、それらをできるだけ独立な2組の素性集合に分類する方法を考案することが大事であろう。

Co-training で素性集合に独立性を仮定するのは、概略、事例を追加することで作成できた訓練データの事例の分布が全体の分布と同じであることを保証するためである。ここで全体の分布は予め分かっているので、どのような事例を追加すれば訓練データの事例の分布が、全体の分布に近づくのかを調べることは可能である。しかし問題は、信頼性のあるラベルが付与できた事例でなければ、追加できないことである。つまり全体の分布との整合性、追加事例の信頼性の2点を勘案しながら追加事例を選択してゆく必要がある。Co-training では素性集合の独立性の仮定があるので、ラベルの信頼性の高い事例を単純に追加しただけで、分布の整合性を満足する。本論文では共起性が高いものを過度に追加するのは、分布の整合性が悪くなるためにそれを避けていると見なせる。また本論文では、ある追加候補の事例

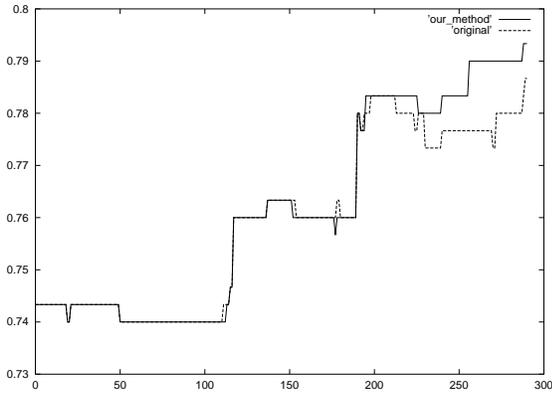


図 2: 正解率の推移 (「声」)

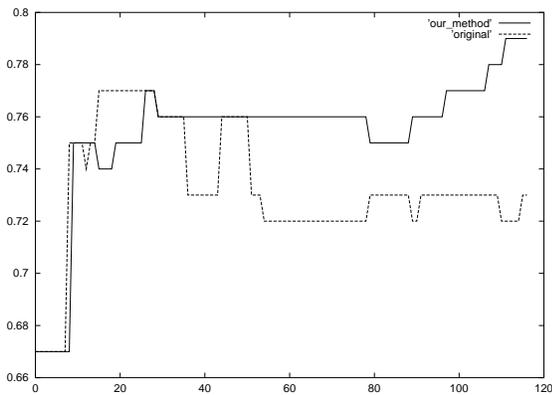


図 3: 正解率の推移 (「壁」)

表 3: 正解率の最大値

	baseline	通常の Co-train	本手法の Co-train
「声」	0.636	0.774	0.824
「トップ」	0.743	0.787	0.793
「壁」	0.670	0.770	0.790

を共起性の条件から追加しなかった場合、次点の候補を追加する。このとき次点の候補につけられたラベルの信頼性は共起性の条件から追加できなかった事例につけられたラベルの信頼性よりも低い。このためこのラベルが誤る確率が多少高くなる。当然、誤ったラベルをつけた事例が  $L$  に追加された場合、学習される規則の精度は徐々に下がってゆく。つまり本手法は追加事例の信頼性を多少犠牲にして、分布の整合性のアドバンテージを取った手法とも見なせる。本論文では共起性の閾値を経験的に 0.3 で固定したが、全体の分布は分っているので、事例ごとに追加すべきかどうかを別の閾値や別の尺度で判断することは可能だと思われる。この点の考察を今後の課題とする。

ラベル付きデータとラベルなしデータを相補的に利用して分類器の精度を高めてゆく手法として、Co-training の他に Naive Bayes と EM アルゴリズムを組み合わせたもの [7] や Transductive 法 [6] が提案されている。これらの手法は独立な 2 つの素性集合を設定する必要がないという長所がある。しかし EM アルゴリズムを用いる手法はデータの発生源に混合モデルを仮定しているために、文書分類以外の自然言語処理の問題へ応用できるのかどうかは明らかではない。またこの手法と Co-training の手法を比較した結果、Co-training が EM アルゴリズムを用いる手法よりも優秀であったという報告 [1] もある。Transductive 法は計算コストが大きすぎ、ラベル付けされていない事例が非常に豊富にある問題に対しては、その実用性に疑問がある。つまり、ある程度独立な 2 つの素性集合を見つけることができれば、Co-training は非常に有効かつ現実的な手法といえる。そのために、独立性が低い 2 つの素性集合から Co-training のような学習をどのように行えば良いかという本論文の研究は重要である。一方、2 つの独立な素性集合は Bagging [3] における複数の分類器で代用できる可能性もある。Bagging を利用して、追加事例とそのクラスを選択する手法も提案されている [8]。更に Co-training は Boosting [5] や “Query by Committee” [9] などの能動学習法と深く関連していると予想している。これらの関係も今後明らかにしてゆきたい。

## 7 おわりに

Co-training は少量のラベル付き訓練データと大量のラベルなし訓練データから分類問題を解決する有効な手法である。しかし Co-training は独立性をもつ 2 つの素性集合を仮定する。この仮定は厳しすぎるため、実際には学習できる規則の精度は頭打ちになる。本論文ではこの原因について概観し、その打開策として追加する事例の選択において、素性間の共起性を調べた。共起性がある程度高い場合には、その事例は追加せずに、次候補の事例を追加することを行った。

実験では「声」、「トップ」、「壁」の 3 つの単語を対象に、語義判別規則の学習に、Co-training を用いた。通常の Co-training に提案手法を組み合わせることで、分類精度を大きく向上させることができた。

追加すべき事例の選択の際に、事例ごとに共起性の閾値を可変にしたり、別種の尺度を考案すること、更に、Co-training と boosting や能動学習法との関係を考察することを今後の課題とする。

## 参考文献

- [1] Kamal Bigam and Rayid Ghani. Analyzing the effectiveness and applicability of co-training. In *9th International Conference on Information and Knowledge Management*, pp. 86–93, 2000.
- [2] Avrim Blum and Tom Mitchell. Combining Labeled and Unlabeled Data with Co-Training. In *11th Annual Conference on Computational Learning Theory (COLT-98)*, pp. 92–100, 1998.
- [3] Leo Breiman. Bagging predictors. *Machine Learning*, Vol. 24, No. 2, pp. 123–140, 1996.
- [4] Michael Collins and Yoram Singer. Unsupervised Models for Named Entity Classification. In *1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pp. 100–110, 1999.
- [5] Yoav Freund, Robert Schapire (訳: 安倍直樹). ブースティング入門. 人工知能学会誌, Vol. 14, No. 5, pp. 771–780, 1999.
- [6] Thorsten Joachims. Transductive inference for text classification using support vector machines. In *16th International Conference on Machine Learning (ICML-99)*, pp. 200–209, 1999.
- [7] Kamal Nigam, Andrew McCallum, Sebastian Thrun, and Tom Mitchell. Text classification from labeled and unlabeled documents using em. In *Machine Learning*, Vol. 39, pp. 103–134, 2000.

- [8] Seong-Bae Park, Byoung-Tak Zhang, and Yung Taek Kim. Word sense disambiguation by learning from unlabeled data. In *38th Annual Meeting of the Association for Computational Linguistics (ACL-00)*, pp. 547–554, 2000.
- [9] H. S. Seung, M. Oppen, and H. Sompolinsky. Query by committee. In *5th annual workshop on Computational Learning Theory (COLT-92)*, pp. 287–294, 1992.
- [10] David Yarowsky. Decision lists for lexical ambiguity resolution: Application to accent restoration in spanish and french. In *32th Annual Meeting of the Association for Computational Linguistics (ACL-94)*, pp. 88–95, 1994.
- [11] David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *33th Annual Meeting of the Association for Computational Linguistics (ACL-95)*, pp. 189–196, 1995.