

複数の構成要素データを扱う多クラス分類器の半教師あり学習法

藤野 昭典 上田 修功 斉藤 和巳

日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

概要: テキストとリンク等を含む Web ページのように、複数の構成要素から成るデータの多クラス単一ラベル分類問題に対して、生成モデル・識別の両アプローチのハイブリッドに基づく分類器の半教師あり学習法を提案する。提案法では、それぞれの構成要素のために、個別に生成モデルを設計するとともに、ラベルありデータが少数であることに起因する生成モデルの学習の偏りの影響を緩和するためのモデルを導入する。さらに、最大エントロピー原理に基づいてこれらのモデルを統合することにより分類器を構築する。リンク等の付加情報を含むテキストデータの分類実験により、生成モデル、識別の各アプローチで類似の分類性能が得られる場合に、提案法では両アプローチより高い分類性能を得られることを確認した。

Semi-supervised Learning of Multi-class Classifiers for Multi-component Data

Akinori FUJINO Naonori UEDA Kazumi SAITO

NTT Communication Science Laboratories, NTT Corporation

Abstract: We present a method for semi-supervised learning of multi-class and single-label classifiers for multi-component data such as web pages consisting of text and links, based on a hybrid generative/discriminative approach. In our formulation, for each component, we design an individual generative model and introduce a model to reduce the effect of the bias associated with the generative model trained on few labeled samples. Then, we construct our classifier by combining these models based on the maximum entropy principle. In our experimental results for text classification using additional information such as links, we confirmed that our classifier outperformed generative and discriminative classifiers especially when the performance of the generative and discriminative classifiers was comparable.

1 はじめに

テキスト、ハイパーリンク、画像などの要素から構成される Web ページのように、複数の構成要素を含むデータ (Multi-Component Data, MCD) の分類問題では、主要な情報が高精度な分類器を設計する上で最も重要な役割を果たすとともに、その他の付加情報についても分類精度の向上に寄与する可能性がある。このため、従来より、Web ページのテキストとハイパーリンク [3, 8]、文書のテキストと引用 [8]、音楽情報とテキスト [2] など、複数の構成要素を扱う分類器が提案されてきた。

教師あり学習の枠組では、確率モデルアプローチによる任意の複数の構成要素を扱う分類器 (以下、MCD 分類器と呼ぶ) として、生成モデル、識別の各アプローチと、両アプローチのハイブリッドに基づく分類器 (生成分類器、識別分類器、ハイブリッド分類器) が提案されてきた。生成、識別分類器と比較してハイブリッド分類器の性能は平均的に高いことが実験的に確認されている [11]。

一方、汎化性能の高い分類器を獲得するためには、膨大なラベルありデータを必要とする場合が多い。しかし、ラベルありデータの作成は、人手によるラベルの付与を必要とし、高コストである。それに対して、ラベルなしデータは比較的容易に収集できる。このため、少数のラベルありデータに加えて多数のラベルなしデータを用いて分類器の汎化性能を向上させる様々な半教師あり学習法が提案されてきた (文献 [12] 参照)。

以上のように、MCD 分類器の設計と半教師あり学習の問題は各分野で別々に研究されてきたが、2つの問題を同時に扱うことで分類器の高精度化を期待できる。そこで、本稿では、MCD 分類器の半教師あり学習の問題に対して、生成モデル・識別両アプローチのハイブリッドに基づく分類器設計法を提案する。提案法は、ラベルありデータで学習される各構成要素の生成モデルと、ラベルなしデータを分類器の学習に効果的に用いるために筆者らが最近考案した偏り補正モデル [5] を構成要素ごとに導入し、これらのモデルを最大エントロピー (ME) 原理 [1] に基づいて統合することの特徴とする。

他の手法として、MCD を扱う生成、識別分類器に、従来の生成モデル、識別の各アプローチに基づく半教師あり学習法 [10, 6] を各々適用することが考えられる。3つのテストコレクションを用いた MCD の分類実験により、生成モデル、識別アプローチに基づく半教師あり学習法の応用と比べて、提案法が汎化性能の高い分類器を獲得するのに有用であることを確認する。

2 MCD 分類器への従来の半教師あり学習法の適用

本稿では、複数の構成要素から成るデータ (MCD) の多クラス単一ラベル分類問題に対して、半教師あり学習に基づいて分類器を設計することを課題とする。 J 個の構成要素から成るデータは、構成要素 j の特徴ベクトル x^j を用いて、 $x = \{x^1, \dots, x^j, \dots, x^J\}$

で表される。このデータ \mathbf{x} に対して、 K 個の候補の中から1つのクラス $y \in \{1, \dots, k, \dots, K\}$ を選択する分類器を設計する。その際、分類器の学習は、ラベルありデータ集合 $D_l = \{\mathbf{x}_n, y_n\}_{n=1}^N$ とラベルなしデータ集合 $D_u = \{\mathbf{x}_m\}_{m=1}^M$ ($M \gg N$) から成る訓練データ集合 $D = \{D_l, D_u\}$ を用いて行う。

以下に、MCD を扱う生成、識別分類器に、従来の生成モデル、識別の各アプローチに基づく半教師あり学習法 [10, 6] を適用する方法を述べる。

2.1 生成分類器の半教師あり学習

生成分類器は、データ \mathbf{x} とクラスラベル y の同時確率密度 $p(\mathbf{x}, y)$ のモデルを用いて設計される。同時確率密度は、各々のクラスで各構成要素は“独立”に生成されると仮定し、構成要素ごとに設計される生成モデル (以下、構成要素モデルと呼ぶ) $p(\mathbf{x}^j|k; \theta_k^j)$ を用いて $p(\mathbf{x}, k; \Theta) = P(k) \prod_{j=1}^J p(\mathbf{x}^j|k; \theta_k^j)$ のようにモデル化される [2]。ここで、 $\Theta = \{\theta_k^j\}_{j,k}$ はクラス k における構成要素 j のモデルパラメータの集合を表す。訓練データを用いて Θ の推定値 $\hat{\Theta} = \{\hat{\theta}_k^j\}_{j,k}$ を求めた後、データのクラス事後確率を、ベイズ則により、

$$P(y = k|\mathbf{x}; \hat{\Theta}) = \frac{P(k) \prod_{j=1}^J p(\mathbf{x}^j|k; \hat{\theta}_k^j)}{\sum_{k'=1}^K P(k') \prod_{j=1}^J p(\mathbf{x}^j|k'; \hat{\theta}_{k'}^j)} \quad (1)$$

で与える。生成分類器は、データが属するクラスがクラス事後確率を最大にする k であると推定する。

生成分類器の半教師あり学習は、ラベルなしデータに混合モデルを仮定することで実現される [10]。すなわち、ラベルなしデータの確率密度を、 $p(\mathbf{x}; \Theta) = \sum_{k=1}^K p(\mathbf{x}, k; \Theta)$ で表される分布でモデル化する。そして、訓練データ集合 D が与えられた下で、 Θ の事後確率密度 $p(\Theta|D)$ を最大化する値をパラメータ Θ の推定値として求める (MAP 推定)。即ち、ベイズ則により、 $\log p(\Theta|D)$ に相当する目的関数:

$$F(\Theta) = \sum_{n=1}^N \log P(y_n) \prod_{j=1}^J p(\mathbf{x}_n^j|y_n; \theta_{y_n}^j) + \sum_{m=1}^M \log \sum_{k=1}^K P(k) \prod_{j=1}^J p(\mathbf{x}_m^j|k; \theta_k^j) + \log p(\Theta) \quad (2)$$

を最大化する Θ を求める。ここで、 $p(\Theta)$ は Θ の事前確率分布を表す。 $F(\Theta)$ を最大化する Θ は EM アルゴリズム [4] を用いて求められる。

Θ の推定は、 N と M の大きさに影響される。 $M \gg N$ のとき、 Θ の推定は、式 (2) の第二項に強く影響され、ラベルなしデータによる教師なし学習に近い推定となる。それ故、混合モデルの仮定が実データに対して適切でない場合、ラベルなしデータをパラメータ学習に用いることで分類精度がむしろ悪化する危険性がある。この問題に対処すべく、ラベルなしデータのパラメータ学習への寄与を調整する重みパラメータ λ を導入する方法 (EM- λ) が提案されている [10]。EM- λ の目的関数は、式 (2) の第二項に $\lambda \in [0, 1]$ を乗することで得られる。

2.2 識別分類器の半教師あり学習

識別分類器は、構成要素ごとにデータを分離せずに、データのクラス事後確率 $P(k|\mathbf{x})$ を直接モデル化して構築できる。例えば、多項ロジスティック回帰 (MLR) モデル [7] を適用して、以下のように未知パラメータ集合 $W = \{w_k^j\}_{j,k}$ を用いてクラス事後確率分布をモデル化できる。

$$P(y = k|\mathbf{x}; W) = \frac{\exp\left(\sum_{j=1}^J w_k^j \cdot \mathbf{x}^j\right)}{\sum_{k'=1}^K \exp\left(\sum_{j=1}^J w_{k'}^j \cdot \mathbf{x}^j\right)} \quad (3)$$

ここで、 $w_k^j \cdot \mathbf{x}^j$ は w_k^j と \mathbf{x}^j の内積を表す。

MLR の半教師あり学習を実現する一手法として、最小エントロピー正則化項 (MER) [6] が提案されている。この方法では、ラベルなしデータのクラス事後確率のエントロピーをクラス分離度の尺度として用い、このエントロピーを最小化することでラベルなしデータをよく分離するように MLR を学習する。MER を用いた MLR の学習では、以下の目的関数を最大化する W をパラメータの推定値とする。

$$F(W) = \sum_{n=1}^N \log P(y_n|\mathbf{x}_n; W) + \lambda \sum_{m=1}^M \sum_{k=1}^K P(k|\mathbf{x}_m; W) \log P(k|\mathbf{x}_m; W) + \log p(W) \quad (4)$$

ここで、 λ は重みパラメータであり、 $p(W)$ は W の事前確率分布を表す。

3 提案法

MCD 分類器の半教師あり学習の問題に対して、生成モデル・識別の両アプローチのハイブリッドに基づく分類器設計法を提案する。提案法では、生成モデルと識別学習の双方の利点を活かした半教師あり学習を実現するために、最近、我々が考案した偏り補正モデル (Bias Correction Model, BCM) [5] を導入して MCD 分類器を構築する。本論文では、便宜上、提案法を H-BCM (Hybrid with BCMs) と呼ぶ。

H-BCM では、まず、クラスごとに構成要素モデル $p(\mathbf{x}^j|k; \theta_k^j)$ を設計し、ラベルありデータ集合 D_l を用いて構成要素モデルを個々に学習する。ラベルありデータ数が少数の場合、パラメータの推定値 $\hat{\Theta} = \{\hat{\theta}_k^j\}_{j,k}$ は、真のデータ分布の最適な近似を与えるパラメータ値から大きく外れる危険性がある。すなわち、少数のラベルありデータで学習される構成要素モデルは、しばしば高い偏りをもつ。この偏りの影響を緩和するため、各クラスの構成要素ごとに偏り補正モデル $p(\mathbf{x}^j|k; \psi_k^j)$ を導入する。偏り補正モデルには、構成要素モデルと同型の分布 (パラメータ $\Psi = \{\psi_k^j\}_{j,k}$ は $\hat{\Theta}$ と異なる) 分布を与える。そして、最大エントロピー (ME) 原理による構成要素モデルと偏り補正モデルの統合に基づいて分類器を構築する。すなわち、H-BCM は、生成過程に基

づくデータの特徴を与える生成モデルの長所と、識別率を最大化する識別学習の長所を兼ね備えたハイブリッドとなっている。

具体的には、分類器の識別関数を、構成要素モデルと偏り補正モデルに関する制約を満たす、エントロピー基準の下で最も一様なクラス事後確率分布 (ME モデル) $R(k|\mathbf{x})$ により定義する。 $R(k|\mathbf{x})$ に生成モデル $p(\mathbf{x}|k; \theta_k)$ の特性を反映させるため、ラベルありデータの経験分布 $\hat{p}(\mathbf{x}, k) = \sum_{n=1}^N I_{\mathbf{x}_n}(\mathbf{x}) I_{y_n}(k) / N$ による生成モデルの対数尤度の期待値と、 $R(k|\mathbf{x})$ による生成モデルの対数尤度の期待値が等しい、という制約を与える。この制約は、 \mathbf{x} の経験分布 $\hat{p}(\mathbf{x}) = \sum_{n=1}^N I_{\mathbf{x}_n}(\mathbf{x}) / N$ を用いて、以下の式で表される。

$$\sum_{\mathbf{x}, k} \hat{p}(\mathbf{x}, k) \log p(\mathbf{x}^j | k; \hat{\theta}_k^j) \\ = \sum_{\mathbf{x}, k} \hat{p}(\mathbf{x}) R(k|\mathbf{x}) \log p(\mathbf{x}^j | k; \hat{\theta}_k^j), \forall j \quad (5)$$

ここで、 $I_{\mathbf{x}_n}(\mathbf{x})$ は、 $\mathbf{x} = \mathbf{x}_n$ のときに 1、それ以外の場合に 0 である指示関数を表す。偏り補正モデル $p(\mathbf{x}|k; \psi_k)$ に関する制約もまた、式 (5) と同様の形で与える。さらに、データが帰属するクラスの偏りを ME モデルに反映させるために、 $R(k|\mathbf{x})$ によるクラス確率の推定値と、ラベルありデータの経験分布によるクラス確率の推定値が等しい、という制約を与える。この制約は以下の式で表すことができる。

$$\sum_{\mathbf{x}} \hat{p}(\mathbf{x}, k) = \sum_{\mathbf{x}} \hat{p}(\mathbf{x}) R(k|\mathbf{x}), \forall k \quad (6)$$

以上の制約の下で、クラス事後確率分布のエントロピー $H(R) = -\sum_{\mathbf{x}, k} \hat{p}(\mathbf{x}) R(k|\mathbf{x}) \log R(k|\mathbf{x})$ を最大化することにより、生成モデル、偏り補正モデルとクラスの偏りを反映した ME モデルのクラス事後確率分布:

$$R(y = k|\mathbf{x}; \hat{\theta}, \Psi, \Gamma) \\ = \frac{1}{Z} e^{\mu_k} \prod_{j=1}^J p(\mathbf{x}^j | k; \hat{\theta}_k^j)^{\gamma_{1j}} p(\mathbf{x}^j | k; \psi_k^j)^{\gamma_{2j}} \quad (7)$$

を導出できる。ここで、 $Z = \sum_{k=1}^K R(k|\mathbf{x}; \hat{\theta}, \Psi, \Gamma) = 1$ とするための正規化項である。また、 $\Gamma = \{\{\gamma_{1j}, \gamma_{2j}\}_{j=1}^J, \{\mu_k\}_{k=1}^K\}$ はラグランジュ乗数である。 γ_{1j} と γ_{2j} は構成要素モデルと偏り補正モデルの統合の重みを、 μ_k はクラス k の出現の偏りを与える。H-BCM では、構成要素モデルと偏り補正モデルの ME 原理に基づく統合により得られる $R(k|\mathbf{x}; \hat{\theta}, \Psi, \Gamma)$ を分類器を与えるクラス事後確率分布とする。

偏り補正モデルのパラメータ Ψ は、ラベルありデータが十分に与えられないことに起因する分類器 $R(k|\mathbf{x}; \hat{\theta}, \Psi, \Gamma)$ の学習の偏りを緩和することを目的として、ラベルなしデータを用いて学習される。具体的には、 Γ の値が与えられるとき、分類器がラベルなしデータに与える識別関数の値の総和に基づく目的関数 $F(\Psi|\Gamma) = \sum_{m=1}^M \log \sum_{k=1}^K \{e^{\mu_k} \prod_{j=1}^J p(\mathbf{x}_m^j | k; \hat{\theta}_k^j)^{\gamma_{1j}} p(\mathbf{x}_m^j | k; \psi_k^j)^{\gamma_{2j}}\} + p(\Psi)$ の最大化により Ψ を推定する。しかし、 Γ は推定すべき未知パラメータであり、 Ψ と互いに依存関係がある。このため、 Ψ と Γ を交互かつ反復的に学習する [5]。

4 評価実験

4.1 テストコレクション

評価実験には、テキスト分類問題のベンチマークテストにしばしば用いられる WebKB¹ と Cora²、20newsgroups (20news)³ の 3 つのテストコレクションを用いた。

WebKB は大学の Web ページを集めたものであり、7 つのカテゴリに分類されている。実験には、 *student, faculty, course, project* の 4 つのカテゴリに含まれる 4199 の Web ページを用いた。Web ページの特徴ベクトルは、本文 (MT)、他のページへのリンク (OL)、他のページからのリンク (IL)、アンカーテキスト (AT) の 4 つの構成要素を抽出して作成した。IL はこのテストコレクションに含まれる Web ページのみから抽出し、AT として当該ページをリンクしている他のページが参照に用いているアンカーテキストを抽出した。MT と AT では単語の出現頻度、IL と OL では URL の出現頻度により構成要素の特徴ベクトルを作成した。

Cora は 3 万以上の技術論文の概要と引用情報を集めたものであり、70 カテゴリのいつれかに分類されている。実験には、 */ArtificialIntelligence/MachineLearning/** の 7 つのカテゴリに属する 4240 論文を用いた。論文の特徴ベクトルは、本文 (MT)、著者名 (AU)、引用論文 (CI) の 3 つの構成要素を抽出して作成した。MT では単語の出現頻度を、AU と CI ではそれぞれ著者名と引用論文の出現を表す特徴ベクトルを作成した。

20news は、UseNet の記事を集めたものであり、20 グループに分類されている。実験には、20 グループのうち、 *comp.** の 5 つのグループに属する 4881 記事を用いた。構成要素として、各記事から “Subject:” のあとに続くタイトル (T) とそれ以外の本文 (M) を抽出した。M と T の特徴ベクトルは単語の出現頻度により作成した。

4.2 実験方法

H-BCM の性能を評価するため、2 節で述べた従来の半教師あり学習法の応用に基づく 2 つの手法 (EM- λ , MLR/MER) との比較実験を行った。実験では、H-BCM の構成要素モデルと偏り補正モデルにナイーブベイズ (NB) モデル [10] を適用した。また、EM- λ にも構成要素モデルとして NB モデルを用いた。

EM- λ と MLR/MER では、ラベルなしデータのモデルパラメータ学習への寄与度を表す重みパラメータ λ を設定する必要がある。本実験では、EM- λ の重みパラメータの候補として、 $\{0.01, 0.05, 0.1, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8, 0.9, 1.0\}$ の 14 種類の値を用いた。パラメータ値は、文献 [10] で提案された方法に従って、ラベルありデータの leave-one-out 交差検定法による分類精度が最も高くなるパラメータ値を選択した。また、MLR/MER の重みパラメータ λ は $\{(0.1 \times 10^{-n}, 0.2 \times 10^{-n}, 0.5 \times 10^{-n})_{n=0}^4, 1\}$ の 16 候補から選択した。パラメータ値は、EM- λ と同様に交差検定法を用いて訓練データのみから決定

¹ <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/webkb-data.tar.gz>

² <http://www.cs.umass.edu/mccallum/data/cora-classify.tar.gz>

³ <http://people.csail.mit.edu/jrennie/20NewsGroups/20news-18828.tar.gz>

表 1: 異なるラベルありデータ数 N に対する H-BCM, EM- λ , MLR/MER の分類精度 (%)

(a) WebKB ($M = 2500, K = 4$)				
N	N/M	H-BCM	EM- λ	MLR/MER
32	0.0128	69.9 (5.1)	73.5 (3.1)	63.5 (5.0)
64	0.0256	76.2 (2.8)	76.0 (2.5)	72.2 (2.7)
128	0.0512	81.8 (1.7)	77.7 (2.4)	78.5 (2.2)
256	0.1024	85.0 (1.3)	80.6 (1.7)	84.3 (1.7)
512	0.2048	87.2 (1.1)	82.7 (1.2)	88.7 (1.3)

(b) Cora ($M = 2000, K = 7$)				
N	N/M	H-BCM	EM- λ	MLR/MER
56	0.028	76.4 (3.6)	71.6 (4.2)	55.7 (4.2)
112	0.056	83.0 (2.1)	77.9 (1.6)	63.8 (2.3)
224	0.112	85.7 (1.2)	81.6 (1.1)	72.9 (1.1)
448	0.224	87.4 (1.3)	83.8 (1.0)	78.6 (1.5)
896	0.448	89.1 (0.7)	86.8 (1.1)	82.7 (1.2)

(c) 20news ($M = 2500, K = 5$)				
N	N/M	H-BCM	EM- λ	MLR/MER
40	0.016	64.8 (3.0)	53.6 (6.6)	50.2 (5.3)
80	0.032	71.6 (1.8)	59.4 (6.4)	57.2 (3.3)
160	0.064	75.7 (1.5)	66.3 (4.6)	65.3 (3.2)
320	0.128	78.5 (1.4)	71.3 (3.2)	72.3 (1.2)
640	0.256	82.0 (1.2)	76.0 (1.5)	77.7 (1.1)
1280	0.512	85.0 (1.0)	80.1 (1.8)	81.0 (1.2)

[6] すべきであるが、パラメータ学習の計算コストが高いため、テストデータの分類精度を最も高くするパラメータ値を選択して性能比較を行った。

各手法の性能評価は、各テストコレクションからランダムにデータを選択して 10 種類の訓練・テストデータセットを作成し、10 回の実験で得られた分類精度の平均値を求めて行った。評価に用いたテストデータ数は各コレクションで 1000、訓練に用いたラベルなしデータ数 M は 2500 (WebKB), 2000 (Cora), 2500 (20news) とした。

4.3 実験結果と考察

表 1 に、各テストコレクションで、ラベルありデータ数 N を変えて実験を行ったときの各手法の分類精度の平均値を示す。括弧内の数値は標準偏差を示す。

H-BCM, EM- λ , MLR/MER の 3 つの手法の分類精度を調べた結果、WebKB で EM- λ と MLR/MER の分類精度の差が大きな場合を除いて、H-BCM の性能が他の手法を上回っていた。この結果より、MCD の分類問題において、ラベルなしデータを学習に用いた分類器の汎化性能を向上させるのに H-BCM が有用であるといえる。

EM- λ と比較して、H-BCM は、WebKB でラベルありデータ数が少数の場合を除いて高い性能を示した。一般的に、大量のラベルありデータが与えられる場合、識別アプローチは生成モデルアプローチよりも高い分類精度を与えることが知られている [9]。この実験結果は、 λ の選択を識別的に行うことを除いて生成モデルアプローチの枠組で分類器を学習する EM- λ には、高精度な分類を達成できない本質的な限界があることを示唆している。

H-BCM の性能は、MLR/MER と比較して、WebKB でラベルありデータ数が多数の場合を除いて高かった。この結果は、MLR/MER には少数のラベルありデータに対して過学習する傾向がある一方、H-BCM では、過学習を抑える生成モデルの特徴を有しているため、MLR/MER よりも高い分類精度が得られたと考えられる。過学習を引き起こさない

ように十分なラベルありデータが与えられるときに、識別アプローチは提案法よりも高い分類精度を与えられられる。

5 まとめ

複数の構成要素を含むデータの多クラス単一ラベル分類問題に対して、生成モデル・識別の両アプローチのハイブリッドに基づく分類器の半教師あり学習法を提案した。提案法は、各構成要素に対して、個々に設計される生成モデルと、ラベルありデータが少数であることに起因する学習の偏りを緩和するために導入されるモデルを、最大エントロピー原理に基づいて統合することを特徴とする。3 つのテストコレクションを用いた分類実験により、生成モデル、識別の各アプローチで類似の分類精度が得られる場合にとくに、提案法では両アプローチよりも高い分類精度を得られることを確認した。今後の課題は、テキストと画像といった異なる分布型でモデル化される異種情報データに対して提案法の有用性を確認することである。

参考文献

- [1] Berger, A. L., Della Pietra, S. A. and Della Pietra, V. J.: A maximum entropy approach to natural language processing, *Computational Linguistics*, Vol. 22, No. 1, pp. 39–71 (1996).
- [2] Brochu, E. and Freitas, N.: “Name that song!”: A probabilistic approach to querying on music and text, *Advances in Neural Information Processing Systems 15*, MIT Press, Cambridge, MA, pp. 1505–1512 (2003).
- [3] Chakrabarti, S., Dom, B. and Indyk, P.: Enhanced hypertext categorization using hyperlinks, *Proceedings of ACM International Conference on Management of Data (SIGMOD-98)*, pp. 307–318 (1998).
- [4] Dempster, A. P., Laird, N. M. and Rubin, D. B.: Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*, Vol. 39, pp. 1–38 (1977).
- [5] 藤野昭典, 上田修功, 齊藤和巳: 半教師あり学習のための生成・識別ハイブリッド分類器の設計法, *人工知能学会論文誌*, Vol. 21, No. 3, pp. 301–309 (2006).
- [6] Grandvalet, Y. and Bengio, Y.: Semi-supervised learning by entropy minimization, *Advances in Neural Information Processing Systems 17*, MIT Press, Cambridge, MA, pp. 529–536 (2005).
- [7] Hastie, T., Tibshirani, R. and Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, New York Berlin Heidelberg (2001).
- [8] Lu, Q. and Getoor, L.: Link-based text classification, *IJCAI Workshop on Text-Mining & Link-Analysis (TextLink 2003)* (2003).
- [9] Ng, A. Y. and Jordan, M. I.: On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes, *Advances in Neural Information Processing Systems 14*, MIT Press, Cambridge, MA, pp. 841–848 (2002).
- [10] Nigam, K., McCallum, A., Thrun, S. and Mitchell, T.: Text classification from labeled and unlabeled documents using EM, *Machine Learning*, Vol. 39, pp. 103–134 (2000).
- [11] Raina, R., Shen, Y., Ng, A. Y. and McCallum, A.: Classification with hybrid generative/discriminative models, *Advances in Neural Information Processing Systems 16*, MIT Press, Cambridge, MA (2004).
- [12] Seeger, M.: Learning with labeled and unlabeled data, Technical report, University of Edinburgh (2001).