

生成・識別モデルの統合に基づく半教師あり学習法と その多重分類への応用

藤野 昭典 上田 修功 磯崎 秀樹

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: 各データが複数のカテゴリに属する多重分類問題に対して、ラベルありデータとラベルなしデータを用いた半教師あり学習により分類器を設計する手法を提案する。提案法では、ラベルありデータで学習させる識別モデルとラベルなしデータで学習させる生成モデルの統合により分類器を得る。提案法を多重テキスト分類問題に適用するため、識別モデルに対数線形モデルを、生成モデルにナイーブベイズモデルを用いる。実テキストデータからなる3つのテストコレクションを用いた実験で、従来の対数線形モデルとナイーブベイズモデルの半教師あり学習法と比較して、提案法ではより高い汎化能力を持つ多重分類器を得られることを確認した。

A Semi-supervised Learning Method based on Generative/Discriminative Model Combination and its Application to Multi-label Classification

Akinori FUJINO Naonori UEDA Hideki ISOZAKI

NTT Communication Science Laboratories, NTT Corporation

Abstract: We propose a method for designing semi-supervised multi-label classifiers, which select one or more category labels for each data example and are trained on labeled and unlabeled examples. The proposed method is based on a combination of discriminative models trained on labeled examples with generative models trained on unlabeled examples. We employed a log-linear model and a naive Bayes model as the discriminative and generative models, respectively, for multi-label text classification problems. Using three test collections consisting of real text data, we confirmed experimentally that the proposed method provided better multi-label classifiers with high generalization ability than conventional semi-supervised learning methods of log-linear and naive Bayes models.

1 はじめに

大量のデータをインターネット等で容易に取得できる現在、効率的にデータを管理するために内容を表すカテゴリラベルをデータに付与する技術のニーズが高まっている。各データに1つ以上のカテゴリラベルを付与する問題は多重分類と呼ばれ、近年、機械学習に基づく多重分類法の研究が行われてきた(e.g.[7])。

機械学習法では、一般的に、属するクラス¹が既知のデータ(ラベルありデータ)を用いて分類器を学習させる。このとき、多数のラベルありデータを用いるほど汎化性能が高い分類器を得やすいことが知られている。しかし、ラベルありデータの作成には対象のデータに精通した専門家によるラベル付けが必要であり、大量に作成するには高いコストを要する。一方、属するクラスが未知のデータ(ラベルなしデータ)を集めるのは比較的容易である。このため、ラベルなしデータをラベルありデータと同時に学習に用いて分類器の性能を向上させる半教師あり学習の研究が近年盛んになっている。

分類器の半教師あり学習法として、生成モデルと識別モデルの各アプローチに基づく手法が提案されている(e.g.[8, 5, 6])。それに対して、我々は、最近、生成モデルと識別学習のハイブリッドに基づく手法を提案し、生成・識別モデルの各アプローチに対す

るハイブリッド法の優位性を実験的に確認した[4]。このハイブリッド法では、ラベルありデータで学習させる生成モデルとラベルなしデータで学習させる生成モデルの重み付き統合により分類器を得る。統合の重みを得るのに識別学習を用いる。

このハイブリッド法は、データに1つのカテゴリラベルが付与される単一ラベル分類問題を対象とした手法であるため、多重分類問題に適用するには、カテゴリラベルごとに付与の可否を判定する2値分類に用いる必要がある。この適用方法では、カテゴリラベルが付与されないデータの生成モデルを設計する必要がある。しかし、対象のカテゴリラベルが付与されないデータにはそれぞれ別の異なるカテゴリラベルが複数個付与されている。また、対象のカテゴリラベルが付与されているデータの中にも別のカテゴリラベルが付与されているものがある。それ故、適切な生成モデルを設計することは容易ではない。ラベルありデータによく適合する生成モデルを与えられなければ、生成モデル同士に統合に基づくハイブリッド法では識別モデルアプローチよりも性能の高い分類器を得られない危険性がある。

そこで本研究では、識別モデルと生成モデルの統合により多重分類器を設計する手法を提案する。本稿では、識別モデルと生成モデルの統合に基づく半教師あり学習法の基本的な枠組と、本手法に基づいて多重テキスト分類器を設計する方法を示す。実テキストデータから成る3つのテストコレクションを用いた評価実験により提案法が高性能な多重分類器を得るのに有効であることを確認する。

¹本稿では、カテゴリラベルの付与の有無で分けられたデータのサブ集合をクラスと呼ぶ。

2 半教師あり学習による多重分類

本稿では、各データに付与すべきカテゴリラベルを、 K 個の候補の中から複数個選択する多重分類問題に焦点を当てる。多重分類問題に対し、データの特徴ベクトル $\mathbf{x}$ から、クラスベクトル $\mathbf{y} = (y_1, \dots, y_K, \dots, y_K)^T$, $y_k \in \{1, 0\}$ ($\mathbf{y} \neq \mathbf{0}$) を推定する分類器を設計する。ここで、 $y_{nk} = 1$ ($y_{nk} = 0$) は n 番目のデータ $\mathbf{x}_n$ に k 番目のカテゴリラベルが付与される (付与されない) ことを表す。分類器の学習には、ラベルありデータ $D_l = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N$ と多数のラベルなしデータ $D_u = \{\mathbf{x}_m\}_{m=1}^M$ ($M \gg N$) から成る訓練データ集合 $D = \{D_l, D_u\}$ を用いる。

3 提案法の枠組

本稿では、生成モデルと識別モデルの統合に基づく半教師あり学習により分類器を設計する手法を提案する。提案法では、ラベルありデータで学習させた識別モデルとラベルなしデータで学習させた生成モデルを用いて分類器の条件付確率モデルを与える。

生成モデルは、データの特徴ベクトル $\mathbf{x}$ とクラスベクトル $\mathbf{y}$ の同時確率分布 $p(\mathbf{x}, \mathbf{y})$ のモデルであり、データの種類に応じて設計される。例えば、テキストデータの生成モデルの設計には多項分布、画像データにはガウス分布などが利用される。提案法では、ラベルなしデータを用いて生成モデル $p(\mathbf{x}, \mathbf{y}; \Theta)$ のパラメータ Θ を学習させる。仮に、ラベルなしデータ $\mathbf{x}_m$ の属するクラス $\mathbf{y}_m$ が既知であれば、パラメータ Θ の事後確率密度 $p(\Theta | D_u)$ に相当する目的関数 $J_g(\Theta) = \sum_{m=1}^M \sum_{\mathbf{y}} I_{\mathbf{y}_m}(\mathbf{y}) \log p(\mathbf{x}_m, \mathbf{y}; \Theta) + \log p(\Theta)$ を最大化させる値を Θ の推定値として求めることができる (MAP 推定)。 $p(\Theta)$ は Θ の事前確率分布であり、 $I_{\mathbf{y}_m}(\mathbf{y})$ は、 $\mathbf{y} = \mathbf{y}_m$ の場合に 1, $\mathbf{y} \neq \mathbf{y}_m$ の場合に 0 の値をとる指示関数である。しかし、 $\mathbf{y}_m$ は未知であるため、 $I_{\mathbf{y}_m}(\mathbf{y})$ の値は定まらない。そこで、 $I_{\mathbf{y}_m}(\mathbf{y})$ の期待値を与える条件付確率分布 $R(\mathbf{y} | \mathbf{x})$ を導入し、 Θ を推定するための目的関数を

$$J(R, \Theta) = \sum_{m=1}^M \sum_{\mathbf{y}} R(\mathbf{y} | \mathbf{x}_m) \log p(\mathbf{x}_m, \mathbf{y}; \Theta) + \log p(\Theta) \quad (1)$$

で与える。但し、 $R(\mathbf{y} | \mathbf{x})$ もまた推定すべき未知の確率分布である。

ラベルなしデータはクラスの情報を持たないため、 $R(\mathbf{y} | \mathbf{x})$ を推定するにはラベルありデータを利用する必要がある。提案法では、 $R(\mathbf{y} | \mathbf{x})$ を推定するのに、ラベルありデータで学習させた識別モデルの識別関数 $f(\mathbf{x}, \mathbf{y}; \hat{W})$ を利用する。ここで、 $\hat{W}$ は識別モデルのパラメータ W の推定値を表す。識別関数は、対数線形モデル等の識別モデルでデータの属するクラスを推定するのに用いられる関数であり、クラスの推定値を $\hat{\mathbf{y}} = \arg \max_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}; \hat{W})$ のように与える。提案法では、 $f(\mathbf{x}, \mathbf{y}; \hat{W})$ と重みパラメータ γ で与えられる条件付確率分布

$$P_{\gamma}(\mathbf{y} | \mathbf{x}; \hat{W}) = \frac{\exp\{\gamma f(\mathbf{x}, \mathbf{y}; \hat{W})\}}{\sum_{\mathbf{y}'} \exp\{\gamma f(\mathbf{x}, \mathbf{y}'; \hat{W})\}} \quad (2)$$

が $R(\mathbf{y} | \mathbf{x})$ の近似を与えると仮定し、両確率分布の KL 距離 $D(R(\mathbf{y} | \mathbf{x}) || P_{\gamma}(\mathbf{y} | \mathbf{x}; \hat{W}))$ の最小化に基づく制約を導入する。そして、この制約の下で、式 (1) で与えた $J(R, \Theta)$ を最大化させる $R(\mathbf{y} | \mathbf{x})$ を推定確率分布とする。すなわち、 $R(\mathbf{y} | \mathbf{x})$ と Θ を推定するための目的関数を

$$J_{\gamma, \beta}(R, \Theta) = \sum_{m=1}^M \sum_{\mathbf{y}} R(\mathbf{y} | \mathbf{x}_m) \log p(\mathbf{x}_m, \mathbf{y}; \Theta) - \frac{1}{\beta} \sum_{m=1}^M \sum_{\mathbf{y}} R(\mathbf{y} | \mathbf{x}_m) \log \frac{R(\mathbf{y} | \mathbf{x}_m)}{P_{\gamma}(\mathbf{y} | \mathbf{x}_m; \hat{W})} + \log p(\Theta) \quad (3)$$

で与える。 β は生成モデル学習の目的関数と KL 距離に基づく制約のバランスを定める正定数である。 $\sum_{\mathbf{y}} R(\mathbf{y} | \mathbf{x}) = 1$ の制約の下でラグランジュ未定乗数法を適用すると、 $\partial J_{\gamma, \beta} / \partial R = 0$ を満たす条件付確率分布

$$R_{\gamma}(\mathbf{y} | \mathbf{x}; \Theta, \hat{W}) = \frac{P_{\gamma}(\mathbf{y} | \mathbf{x}; \hat{W}) p(\mathbf{x}, \mathbf{y}; \Theta)^{\beta}}{Z_{\gamma}(\mathbf{x}; \Theta, \hat{W})} = \frac{\exp\{f(\mathbf{x}, \mathbf{y}; \hat{W})\}^{\gamma} p(\mathbf{x}, \mathbf{y}; \Theta)^{\beta}}{\sum_{\mathbf{y}'} \exp\{f(\mathbf{x}, \mathbf{y}'; \hat{W})\}^{\gamma} p(\mathbf{x}, \mathbf{y}'; \Theta)^{\beta}} \quad (4)$$

が得られる。但し、 $Z_{\gamma}(\mathbf{x}; \Theta, \hat{W}) = \sum_{\mathbf{y}'} P_{\gamma}(\mathbf{y}' | \mathbf{x}; \hat{W}) \times p(\mathbf{x}, \mathbf{y}'; \Theta)^{\beta}$ である。 $\gamma = (\gamma, \beta)^T$ は識別関数と生成モデルの統合の重みを与える。提案法では、上記の $R_{\gamma}(\mathbf{y} | \mathbf{x}; \Theta, \hat{W})$ を分類器の条件付確率モデルとする。

生成モデルのパラメータ Θ の推定値は、式 (3) に式 (4) を代入して得られる目的関数

$$J_{\gamma}(\Theta) = \frac{1}{\beta} \sum_{m=1}^M \log Z_{\gamma}(\mathbf{x}_m; \Theta, \hat{W}) + \log p(\Theta) \quad (5)$$

の最大化に基づいて求められる。 γ の値を設定すると、EM アルゴリズム [1] のような繰り返し計算を行うことで、 $J_{\gamma}(\Theta)$ を初期値近傍で最大化させる Θ を推定できる。

提案法では、 γ の値を、ラベルありデータを用いた条件付確率モデルの尤度最大化学習に基づいて設定する。紙面の都合により、 γ の値の設定法の詳細を省略する。

4 多重テキスト分類への適用法

本節では、生成モデルにナイーブベイズ (NB) モデルを、識別モデルに対数線形モデルを用いて、提案法を多重テキスト分類問題に適用する方法を述べる。以下に、NB モデルと対数線形モデルの概要と、両モデルを用いて設計される条件付確率モデルを示す。

4.1 NB モデル

本適用法では、 k 番目のカテゴリラベルが付与される場合 ($y_k = 1$) と付与されない場合 ($y_k = 0$) に対して、以下のような生成モデルを設計する。

$$p(\mathbf{x}, y_k; \Theta_k) = p(\mathbf{x} | y_k; \Theta_k) \pi_k(y_k) \quad (6)$$

ここで、 $\pi_k(y_k)$ は y_k の確率を表し、 $\sum_{y_k=0}^1 \pi_k(y_k) = 1$ を満たす。 $p(\mathbf{x}|\mathbf{y}_k; \Theta_k)$ は y_k の条件下での $\mathbf{x}$ の確率密度を表す。 $p(\mathbf{x}|\mathbf{y}_k; \Theta_k)$ のモデル化に、NB モデル [8] を用いる。

NB モデルでは、文書に含まれる単語が独立に出現すると仮定し、 i 番目の単語の出現頻度 x_i で与えられる特徴ベクトル $\mathbf{x} = (x_1, \dots, x_i, \dots, x_V)^T$ を用いて文書を表現する。ここで、 V は文書集合全体に含まれる単語の種類 (語彙) の総数を表す。そして、 y_k での文書の確率密度を $p(\mathbf{x}|\mathbf{y}_k; \Theta_k) \propto \prod_{i=1}^V \{\theta_{ki}(y_k)\}^{x_i}$ で表される多項分布でモデル化する。ここで、 $\theta_{ki}(y_k)$ は y_k で i 番目の単語が出現する確率を表し、 $\theta_{ki}(y_k) > 0$ 、 $\sum_{i=1}^V \theta_{ki}(y_k) = 1$ を満たす。 $\Theta_k = [\theta_{ki}(y_k)]_{i,y_k}$ は、訓練データを用いて推定すべきパラメータである。

4.2 対数線形モデル

対数線形 (log linear) モデルでは、カテゴリラベルごとに独立な関数 $f_k(\mathbf{x}, y_k; \mathbf{w}_k) = y_k \mathbf{w}_k^T \mathbf{x}$ の線形和で表される識別関数 $f(\mathbf{x}, \mathbf{y}; W) = \sum_{k=1}^K f_k(\mathbf{x}, y_k; \mathbf{w}_k)$ を用いて、 $\mathbf{x}$ に対する $\mathbf{y}$ の条件付確率分布を

$$P(\mathbf{y}|\mathbf{x}; W) = \frac{\exp \left\{ \sum_{k=1}^K f_k(\mathbf{x}, y_k; \mathbf{w}_k) \right\}}{\sum_{\mathbf{y}'} \exp \left\{ \sum_{k=1}^K f_k(\mathbf{x}, y'_k; \mathbf{w}_k) \right\}} \quad (7)$$

で与える。 $W = [\mathbf{w}_1, \dots, \mathbf{w}_k, \dots, \mathbf{w}_K]^T$ はパラメータ行列を表す。 W の推定値は、ラベルありデータ集合 D_l による以下の目的関数を最大化させる W として求めることができる。

$$J_d(W) = \sum_{n=1}^N \log P(\mathbf{y}_n|\mathbf{x}_n; W) + \log p(W) \quad (8)$$

$p(W)$ は W の事前確率分布を表す。

4.3 半教師あり多重分類器のモデル

3 節で述べた手法に従い、NB モデル $p(\mathbf{x}|\mathbf{y}_k; \Theta_k)$ と対数線形モデルの識別関数 $f_k(\mathbf{x}, y_k; \mathbf{w}_k)$ の統合に基づく多重分類器を以下の目的関数を最大化させる条件付確率分布 $R(\mathbf{y}|\mathbf{x})$ でモデル化する。

$$\begin{aligned} J_\gamma(R, \Theta) &= \sum_{k=1}^K \sum_{m=1}^M \sum_{\mathbf{y} \neq \mathbf{0}} R(\mathbf{y}|\mathbf{x}_m) \log p(\mathbf{x}_m|\mathbf{y}_k; \Theta_k) \pi_k(y_k) \\ &\quad - \frac{1}{\beta} \sum_{m=1}^M \sum_{\mathbf{y} \neq \mathbf{0}} R(\mathbf{y}|\mathbf{x}_m) \log \frac{R(\mathbf{y}|\mathbf{x}_m)}{P_\gamma(\mathbf{y}|\mathbf{x}_m; \hat{W})} \\ &\quad + \log p(\Theta) \end{aligned} \quad (9)$$

但し、 $\Theta = \{\Theta_k\}_{k=1}^K$ である。多重分類は各データに 1 つ以上のカテゴリラベルを付与する問題であるため、 $\mathbf{y} = \mathbf{0}$ を除く $\mathbf{y}$ に対して上記の目的関数を与えていることに注意。 $\sum_{\mathbf{y} \neq \mathbf{0}} R(\mathbf{y}|\mathbf{x}) = 1$ の条件下

でラグランジュ未定乗数法を用いると、多重分類器の条件付確率モデル

$$\begin{aligned} R_\gamma(\mathbf{y}|\mathbf{x}; \hat{W}, \Theta, \mu) &= \frac{\prod_{k=1}^K h_k(\mathbf{x}, y_k; \hat{\mathbf{w}}_k, \Theta_k, \mu_k, \gamma)}{\sum_{\mathbf{y}' \neq \mathbf{0}} \prod_{k=1}^K h_k(\mathbf{x}, y'_k; \hat{\mathbf{w}}_k, \Theta_k, \mu_k, \gamma)} \quad (10) \end{aligned}$$

を得ることができる。但し、

$$\begin{aligned} h_k(\mathbf{x}, y_k; \hat{\mathbf{w}}_k, \Theta_k, \mu_k, \gamma) &= \exp\{f_k(\mathbf{x}, y_k; \hat{\mathbf{w}})\} p(\mathbf{x}|\mathbf{y}_k; \Theta_k)^\beta \exp(y_k)^\mu \quad (11) \end{aligned}$$

かつ $\mu_k = \beta \log \pi_k(y_k = 1) / \pi_k(y_k = 0)$ である。 $\gamma = (\gamma, \beta)^T$ と $\mu = (\mu_1, \dots, \mu_k, \dots, \mu_K)^T$ の値には、ラベルありデータによる条件付確率モデルの尤度最大化学習で得られる推定値を用いる。また、式 (9) に式 (10) を代入して得られる目的関数の最大化に基づいて Θ の値を推定する。

5 実験

5.1 テストコレクション

提案法の分類性能を実データを用いた実験で評価した。実験には、多重テキスト分類のベンチマークテストによく用いられる 3 つのテストコレクション (Reuters-21578, RCV1-2, WIPO-alpha) を用いた。

Reuters-21578² データセット (Reuters) は、135 トピックカテゴリからなる Reuters newswire の英文記事を集めたテストコレクションである。実験には、記事を多く含む 10 トピックカテゴリを用いた。

RCV1-2³ データセットは、103 トピックカテゴリからなる Reuters newswire の英文記事を集めたテストコレクションであり、訓練データセットとテストデータセットから構成されている。実験には、訓練データセット内の多くの記事が属する 20 トピックカテゴリを利用した。

WIPO-alpha データセット (WIPO) は、国際特許分類 (IPC) 体系で分類された特許文書からなり、訓練データセットとテストデータセットから構成されている [3]。実験には、訓練データセット内の多くの文書が属する 20 クラスを利用した。

5.2 実験設定

実験では、提案法の性能を、4 つの半教師あり学習に基づく分類器 (JLL/MCL, LL/MER, NB/DC, NB/EM- λ) 及び教師あり学習に基づく分類器 (LL) の性能と比較した。

JLL/MCL は、同時確率分布の対数線形モデル (joint log linear model, JLL モデル) $p(\mathbf{x}, \mathbf{y})$ を用いて設計される分類器であり、multi-conditional learning (MCL) と呼ばれる手法で JLL モデルを学習させる [2]。この手法では、同時確率分布の周辺化で得られる $p(\mathbf{x})$ と $p(\mathbf{y}|\mathbf{x}) = p(\mathbf{x}, \mathbf{y})/p(\mathbf{x})$ で得られる条件付確率分布を用いて半教師あり学習を行う。

LL/MER は、式 (7) で与える条件付確率分布の対数線形 (LL) モデルに基づく分類器であり、最小エン

²<http://www.daviddlewis.com/resources/testcollections/reuters21578/reuters21578.tar.gz>

³<http://www.daviddlewis.com/resources/testcollections/rcv1/>

表 1: N 個のラベルありデータと 5000 個のラベルありデータで学習させた場合の分類精度 (%)

Test Collection	N	提案法	JLL/MCL	LL/MER	NB/DC	NB/EM- λ	LL
Reuters	80	83.4 (1.9)	74.7 (2.6)	72.4 (1.1)	74.8 (2.0)	73.0 (3.3)	72.1 (1.4)
	320	88.2 (0.8)	82.6 (0.6)	82.8 (0.6)	80.5 (1.4)	78.2 (1.4)	82.5 (0.5)
	1280	90.7 (0.4)	87.1 (0.7)	87.7 (0.5)	83.8 (0.4)	82.3 (0.6)	87.4 (0.4)
RCV1-2	160	31.8 (1.8)	30.9 (1.5)	30.7 (1.7)	29.0 (2.0)	25.3 (3.2)	30.7 (1.7)
	640	45.3 (0.9)	45.5 (1.2)	45.1 (0.8)	39.2 (0.7)	35.4 (1.2)	44.8 (1.0)
	2560	55.4 (0.7)	54.6 (0.9)	54.6 (0.6)	45.9 (0.6)	40.6 (0.8)	54.5 (0.9)
WIPO	160	37.9 (3.0)	18.1 (3.8)	11.9 (2.5)	16.5 (2.0)	9.7 (5.7)	12.1 (2.4)
	640	46.8 (1.4)	28.5 (1.9)	24.8 (1.9)	27.2 (2.3)	21.5 (2.6)	24.8 (1.9)
	2560	53.1 (0.7)	37.6 (1.3)	36.9 (1.1)	34.1 (1.4)	30.4 (1.7)	36.9 (1.1)

トロピー正則化 (minimum entropy regularization, MER) 法で LL モデルを学習 [5] させる。

LL は、式 (7) で与える条件付確率分布の対数線形 (LL) モデルに基づく分類器であり、ラベルありデータのみで学習させる。

NB/DC は、文献 [4] で提案された生成モデルと識別学習のハイブリッドに基づく単一ラベル分類器であり、ラベルありデータで学習させる NB モデルとラベルなしデータで学習させる NB モデルの重み付き統合で与えられる。統合の重みは識別学習により与えられる。本実験では、NB/DC をカテゴリごとにラベルの付与を判定する 2 値分類に用いた。

NB/EM- λ は、NB モデルに基づく単一ラベル分類器であり、EM- λ アルゴリズム [8] を用いて NB モデルを学習させる。本実験では、NB/EM- λ を NB/DC と同様に各カテゴリの 2 値分類に用いた。

本稿では、ラベルあり・なしデータとテストデータをランダムに選択して行う実験を同一条件で 10 回繰り返し、その繰り返し実験で得られる分類精度の平均値を用いて分類器の性能を比較した。分類精度は、属するカテゴリを正確に推定できたテストデータの比率を表す。実験では、5000 個のラベルなしデータと N 個のラベルありデータで学習させた分類器の分類精度を調べた。Reuters と RCV1-2, WIPO の各実験でそれぞれ 2545, 5000, 5000 個のテストデータを用いた。

5.3 実験結果と考察

表 1 に、繰り返し実験で得られた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。

提案法の分類精度は、RCV1-2 の $N = 640$ の場合を除いて、JLL/MCL, LL/MER よりも高かった。JLL/MCL と LL/MER ではラベルあり・なしデータで学習させた対数線形モデルのみを用いて分類器を構築するのに対し、提案法ではラベルなしデータによる分類器の学習に生成モデルを利用する。すなわち、提案法では、データの種類に応じて設計される生成モデルの確率分布をラベルなしデータの分布の事前知識として利用する。この事前知識の利用が、ラベルなしデータを用いて分類器の性能を向上させるのに有効であったと考えられる。

提案法の分類精度は NB/DC よりも高く、NB/DC の分類精度は NB/EM- λ よりも高かった。NB/DC の分類精度は、ラベルありデータが少数の場合を除いて、識別モデルに基づく JLL/MCL と LL/MER よりも低かった。NB/DC は、提案法と同様に複数のモデルの統合により分類器を与えるが、統合するモデルに生成モデルのみを用いる点が提案法と異なる。実験結果は、生成モデルのみの統合に基づく

NB/DC では、高性能な多重分類器を得るのに限界があることを示唆している。

6 まとめ

本稿では、識別モデルと生成モデルの統合に基づく半教師あり学習法により多重分類器を設計する手法を提案し、対数線形モデルとナイーブベイズ (NB) モデルを用いて設計した多重テキスト分類器の性能を調べた。3 つのテストコレクションを用いた評価実験により、生成モデルと識別学習のハイブリッドに基づく単一ラベル分類器を各カテゴリの 2 値分類に利用する手法や、従来の対数線形モデルと NB モデルの半教師あり学習法と比較して、提案法ではより汎化性能の高い多重分類器を得られることを確認した。

参考文献

- [1] 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).
- [2] Druck, G., Pal, C., Zhu, X. and McCallum, A.: Semi-supervised classification with hybrid generative/discriminative methods, *Proceedings of 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07)*, pp. 280–289 (2007).
- [3] Fall, C. J., Törösvári, A., Benzineb, K. and Karetka, G.: Automated categorization in the international patent classification, *ACM SIGIR Forum*, Vol. 37, No. 1, pp. 10–25 (2003).
- [4] Fujino, A., Ueda, N. and Saito, K.: Semi-supervised learning for a hybrid generative/discriminative classifier based on the maximum entropy principle, *IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 30, No. 3, pp. 424–437 (2008).
- [5] 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).
- [6] Joachims, T.: Transductive inference for text classification using support vector machines, *Proceedings of the 16th International Conference on Machine Learning (ICML-99)*, pp. 200–209 (1999).
- [7] McCallum, A. K.: Multi-Label Text Classification with a Mixture Model Trained by EM, *AAAI'99 Workshop on Text Learning* (1999).
- [8] 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).