

Multi-Topic Document Classification Introducing A Label Selection
Method Based On Similarity

AKIYAMA Ryutaro

1 研究の背景と目的

近年、情報技術の発達に伴い、電子的に発行される記事や論文、ECサイト等のユーザレビューなど、多様な形式の電子文書が膨大に存在しており、現在も増加し続けている。そのため、ユーザが膨大な文書から目的の文書を探し出すまでには大きな労力を要する場合がある。ユーザが目的の文書を探す際に、膨大な文書が「スポーツ」や「政治」のような文書の内容を表わすトピックごとに分類されていれば、目的に合致した文書を探し出すことが容易になる。このため、文書をトピックごとに自動分類する技術 [1], [2] の重要性は増加しているといえる。

他方、電子文書には複数のトピックを持つものが数多く存在している。例として、「長嶋茂雄氏の国民栄誉賞受賞」の記事は「スポーツ」と「社会」の2つのトピックを有していると考えられる。一般に、電子文書には1つのトピックをもつ文書と複数のトピックをもつ文書が混在している。このような文書に対し、トピックが未知である新規文書の複数のトピックをすべて予測する手法がマルチトピック文書分類手法であり、その研究が盛んに行われている [3]-[7]。

マルチトピック文書分類手法とは、与えられたトピック集合の中から文書の内容を表わすトピックをすべて選び出す手法を指す。以下では、文書の内容を表わすトピック集合をラベルと呼ぶ。例として、「スポーツ」、「政治」、「社会」、「芸能」などのトピックがあった場合、ラベルは {スポーツ, 社会}, {政治, 社会, 芸能} のようなトピックの集合を指す。

マルチトピック文書分類手法には、2クラス分類問題に分解する手法 [5] や Parametric Mixture Model [6], Latent Dirichlet Allocation を用いた手法 [7] など様々なものが存在するが、本研究では、その中でも高精度な予測が可能な最大マージン原理を用いた Maximum Margin Labeling (以下、MML) [3] に注目する。MML は、正解ラベルが既知の文書に対して、最大マージン原理を用いることで、その文書が正解ラベル以外のラベルに誤分類される可能性を低くするように学習を行う手法である。また、MML では計算量削減のため、隣接ラベルと呼ばれる学習データへの類似度が高いラベルのみを選択して学習に用いることで計算量を削減し、現実的な時間での学習を可能としている。隣接ラベルは、正解ラベルと1つのトピックの有無だけが異なるラベルである。例えば、「スポーツ」、「政治」、「社会」などのトピックがあり、{政治, 社会} が正解ラベルの場合、{スポーツ, 政治, 社会}, {社会}, {政治} が隣接ラベルとなる。トピック数が少ないデータセットに対しては、すべてのラベルを用いた場合と隣接ラベルのみを用いた場合で、同程度の予測精度となることが示されている。

MML では、すべての隣接ラベルを学習に用いるが、これは各隣接ラベルが誤分類に与える影響は同程度と解釈していることになる。しかし、隣接ラベルの中でも、誤分類に与える影響には差があると考えられ、この影響が強

いと考えられるラベルを重視した学習を行うことで、分類精度の向上が期待できる。

そこで本研究では、隣接ラベルの中でも特に誤分類に影響を与えるラベルを、ラベルの類似性を利用することで選択し、そのラベルのみを学習に用いる手法を提案する。あるラベルに対するラベルの類似性を、あるラベルが持つトピックと共起しにくいトピックを持つラベルほど低いと定義する。正解ラベルと類似性が低い隣接ラベルは、正解ラベルがもつトピックと共起しにくいトピックを持っているため、生起する可能性が低い。したがって、正解ラベルと類似性が低い隣接ラベルは特に誤分類に影響を与えると考えられる。さらに、隣接ラベルの中からラベルを選択するため、従来の MML と比較して学習に用いるラベル数が減少し、計算量の削減も期待できる。提案手法の有効性を示すためにベンチマークデータを用いた検証実験を行う。

2 Maximum Margin Labeling

以下では、本研究で注目する Maximum Margin Labeling (MML) [3] について説明を行う。ほとんどのマルチトピック文書分類手法はトピックごとに学習データをもつが、MML はラベルごとに学習データをもつため、「スポーツ, 社会」というトピックの組み合わせでのみ重要な単語のような、複数のトピックの組み合わせに特徴的な属性を予測に利用することができる。また、ラベルごとに学習データをもつことに起因する、各ラベルあたりの学習データ数の減少による過学習の危険性を最大マージン原理を応用することで回避している。

以下、 \mathbf{x}_i は i 番目の文書の単語頻度ベクトルであり、ベクトルの各成分は文書内での各単語の出現頻度とする。 L_i は i 番目の文書のラベルとし、ラベルはトピックの集合を表わす。このとき、ラベルが付与された m 組の学習データは $\{\mathbf{x}_i, L_i\}_{i=1}^m$ と表わされる。

2.1 MML の概要

MML は、ラベルの予測のためにラベル空間というベクトル空間を導入する。図 1 に MML の概要を示す。 R^N , R^M はそれぞれ N 次元, M 次元ユークリッド空間, t はトピックを表わしている。

まず、トピック数が T のとき、ラベル L は T 次元ユークリッド空間 R^T 上の 1 点で表わされるので、以下の条件を満たすような関数 $\phi: R^T \rightarrow R^M$ を用いて、これらのラベルをラベル空間に写像する。

$$\langle \phi(L), \phi(L') \rangle = D(L, L'). \quad (1)$$

$$D(L, L') = \begin{cases} \frac{2|L \cap L'|}{|L| + |L'|} & \text{if } |L| \neq 0 \text{ or } |L'| \neq 0. \\ 1 & \text{otherwise.} \end{cases} \quad (2)$$

$\langle \mathbf{A}, \mathbf{B} \rangle$ は \mathbf{A} と \mathbf{B} の内積を表わし、 $|L|$ はラベル L に含まれるトピック数を表わす。ラベル空間上では、ラベル L は $\phi(L)$ となり、空間上に固定されているようになる。つぎに、学習データを用いて、文書の特徴ベクトル \mathbf{x}_i をラベル空間に写像する \mathbf{W} を、 $\mathbf{W}\mathbf{x}_i$ と正解ラベル $\phi(L_i)$

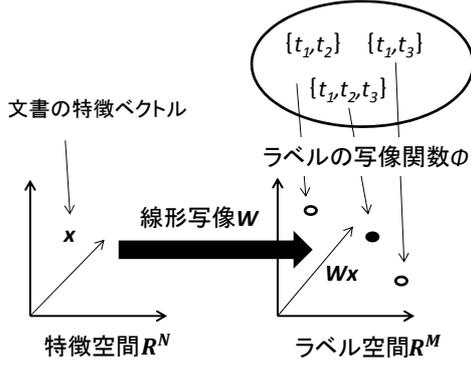


図 1: MML の概要

が近くなるように学習する。そして、ラベルが未知の文書に対して、特徴ベクトル \mathbf{x} を \mathbf{W} でラベル空間に写像し、 $\mathbf{W}\mathbf{x}$ と $\phi(L)$ の内積が最大であるラベル L を文書のラベルとして予測する。

つまり、MML は以下の 3 つのステップから構成される。

1. ラベルの写像
2. 文書の特徴ベクトルの写像 \mathbf{W} の学習
3. ラベルが未知の文書に対するラベルの予測

2.2 写像 \mathbf{W} の学習

ステップ 2 の文書の特徴ベクトルの写像 \mathbf{W} の学習について説明する。MML では、学習データの写像先 $\mathbf{W}\mathbf{x}$ から誤分類される領域へのマージンが最大になるように写像 \mathbf{W} を学習する。ただし、 \mathbf{W} は $M \times N$ 行列であり、 N 次元特徴ベクトルを M 次元ラベル空間に線形写像する行列である。ここで、学習データ $\{\mathbf{x}, L\}$ のマージンは式 (3) で定義される。 λ は正解ラベル以外のラベルを表わしている。

$$\min_{\lambda} \left\langle \mathbf{W}\mathbf{x}, \frac{\phi(L) - \phi(\lambda)}{\|\phi(L) - \phi(\lambda)\|} \right\rangle. \quad (3)$$

一般に、ラベルはトピックの組み合わせ数だけ存在するため、ラベルあたりの学習データ数が減少し、過学習を起こす危険性がある。そこで、マージンが最大になるように学習を行うことで、この問題に対処する。マージン最大化に基づく \mathbf{W} の学習は次の最適化問題として定式化できる。

$$\begin{aligned} \min_{\mathbf{W}} \quad & \frac{1}{2} \|\mathbf{W}\|^2 + C \sum_{i=1}^m \sum_{\lambda \in \Lambda} \xi_i^\lambda, \\ \text{s.t.} \quad & \left\langle \mathbf{W}\mathbf{x}_i, \frac{\phi(L_i) - \phi(\lambda)}{\|\phi(L_i) - \phi(\lambda)\|} \right\rangle \geq 1 - \xi_i^\lambda, \\ & \xi_i^\lambda \geq 0, \\ & \|\mathbf{W}\| = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N W_{ij}^2}. \end{aligned} \quad (4)$$

ただし、 ξ_i^λ は i 番目の文書のラベル λ に対するマージンの制約を破った値 (スラック変数) とする。また、 C はスラック変数に対するペナルティの大きさをコントロールするパラメータである。 Λ はすべての可能なラベルを要素とする集合 (トピック数が T のとき、この集合の要素数は $2^T - 1$ となる) を表わしている。

式 (4) の最適化問題は、ラベルを写像する関数 ϕ を含んでいるため、式 (1), (2) より事前に ϕ を求めておかなければ解くことができない。これに対し、以下の双対問題を定式化することで、計算を容易にする。

$$\begin{aligned} \max_{\{\alpha_i^\lambda: i=1, \dots, m\}} \quad & \sum_{i=1}^m \sum_{\lambda \in \Lambda} \alpha_i^\lambda - \frac{1}{2} \sum_{i,j=1}^m \sum_{\lambda \in \Lambda, \lambda' \in \Lambda} \alpha_i^\lambda \alpha_j^{\lambda'} \\ & \times (\mathbf{x}_i \cdot \mathbf{x}_j) \frac{D(L_i, L_j) - D(L_i, \lambda') - D(\lambda, L_j) + D(\lambda, \lambda')}{2\sqrt{(1 - D(L_i, \lambda))(1 - D(L_j, \lambda'))}}, \\ \text{s.t.} \quad & 0 \leq \alpha_i^\lambda \leq C. \end{aligned} \quad (5)$$

式 (5) では、 ϕ に関する計算はすべて ϕ の内積である D で計算されるため、 ϕ を求める必要がなくなる。

2.2.1 学習に用いるラベルの選択

ラベル数はトピック数に対して指数関数的に増えるため非常に大きく、かつ式 (5) における目的関数はすべてのラベルに関する和を含んでいるため、トピック数が多い場合には膨大な学習時間を要する。これに対し、MML では、学習データ数を減少させるため、一部のラベルのみを学習に用いることで、計算量の削減を図っている。

MML では、マージンが小さいラベルへの誤分類の可能性が高くなる。マージンが小さくなる状況は、ラベル空間上で誤ったラベル λ が正解ラベル L_i と隣接している場合に起こりやすいため、隣接しているラベル (隣接ラベル) のみを選択して学習に用いる。隣接ラベルは、正解ラベルと 1 つのトピックの有無だけ異なるラベルであり、 L_i の隣接ラベル集合 A_i は式 (6) で表わされる。 Δ は対称差を表わす集合の記号である。

$$A_i = \{\lambda \in \Lambda \mid |\lambda \Delta L_i| = 1\}. \quad (6)$$

例えば、「スポーツ」、「政治」、「社会」などのトピックがあり、{政治, 社会} が正解ラベルの場合、{スポーツ, 政治, 社会}, {社会}, {政治} などが隣接ラベルとなる。

また、隣接ラベルのみを学習に用いた場合、式 (7) のように表現できる。

$$\begin{aligned} \max_{\{\alpha_i^\lambda: \lambda \in A_i, i=1, \dots, m\}} \quad & \sum_{i=1}^m \sum_{\lambda \in A_i} \alpha_i^\lambda - \frac{1}{2} \sum_{i,j=1}^m \sum_{\lambda \in A_i, \lambda' \in A_j} \alpha_i^\lambda \alpha_j^{\lambda'} \\ & \times (\mathbf{x}_i \cdot \mathbf{x}_j) \frac{D(L_i, L_j) - D(L_i, \lambda') - D(\lambda, L_j) + D(\lambda, \lambda')}{2\sqrt{(1 - D(L_i, \lambda))(1 - D(L_j, \lambda'))}}, \\ \text{s.t.} \quad & 0 \leq \alpha_i^\lambda \leq C. \end{aligned} \quad (7)$$

式 (7) では、式 (5) の全ラベルに関する和が隣接ラベルに関する和に置き換わっており、計算時間が大きく削減される。また、トピック数が少ないデータセットで実験した結果、すべてのラベルを用いた場合と隣接ラベルのみを用いた場合では、同程度の予測精度となることが示されている [3]。

2.2.2 学習アルゴリズム

式 (7) の学習には、二次計画問題で一般的に使われる解法である最急降下法が用いられる。最急降下法は、関数の傾きが最も急な方向に降下していくことで、関数の極小値を探索するアルゴリズムであり、関数の傾きのみを見るため、単純であり計算も速く行うことができる。以下に、アルゴリズムを示す。

Step1) 全ての α_i^λ の初期値をランダムに決める。

Step2) 全ての (i, λ) について v_i^λ を式 (8) より求める。

$$\begin{aligned} v_i^\lambda = -1 + \sum_{j=1}^m \sum_{\lambda' \in A_j} \alpha_j^{\lambda'} (\mathbf{x}_i \cdot \mathbf{x}_j) \\ \times \frac{D(L_i, L_j) - D(L_i, \lambda') - D(\lambda, L_j) + D(\lambda, \lambda')}{2\sqrt{(1 - D(L_i, \lambda))(1 - D(L_j, \lambda'))}}. \end{aligned} \quad (8)$$

これは目的関数を α_i^λ について偏微分した式である。

Step3) 全ての α_i^λ の値を、式 (9) より更新する。

$$\alpha_i^{\lambda \text{ new}} = \alpha_i^{\lambda \text{ old}} - \varepsilon v_i^\lambda. \quad (9)$$

ε は一回の更新の割合を決めるパラメータである。

Step4) 全ての (i, λ) について、式 (8) の値が 0 になったら終了する。さもなければ、Step2 に戻る。

Step2 より、MML の更新 1 回あたりの計算量は $O(M_A^2)$ となる。ただし、 $M_A = \sum_{i=1}^m |A_i|$ である。

3 提案手法

3.1 提案手法への着眼点

従来の MML は、学習時に正解ラベルと 1 つのトピックの有無だけ異なるラベルである隣接ラベルをすべて用いるため、すべての隣接ラベルが同程度誤分類に影響を与えるとしている。しかし、隣接ラベルの中でも、誤分類に与える影響には差があると考えられ、特に誤分類に影響を与えるラベルを重視した学習を行うことで、分類精度の向上が期待される。そこで本研究では、ラベルの類似性を利用することで、隣接ラベルの中でも特に誤分類に影響を与えるラベルを選択することを考える。

本研究では、あるラベルに対するラベルの類似性を、あるラベルが持つトピックと共起しにくいトピックを持つラベルほど低いと定義する。例えば、「スポーツ」、「政治」、「社会」、「芸能」などのトピックがあり、{政治, 社会} が正解ラベル、「政治」と「スポーツ」は共起することがあるが、「政治」と「芸能」が共起しにくいという場合、ラベル {スポーツ, 政治, 社会} は正解ラベルと類似性が高く、ラベル {政治, 社会, 芸能} は類似性が低いことになる。正解ラベルと類似性が低い隣接ラベルは、正解ラベルがもつトピックと共起しにくいトピックを持っているため、生起する可能性が低い。したがって、正解ラベルと類似性が低い隣接ラベルは特に誤分類に与える影響が大きいと考えられるので、そのラベルのみを選択して学習に用いる方法を提案する。この手法により、特に誤分類に影響を与えるラベルを重視した学習を行うことができ、分類精度の向上が期待できる。また、隣接ラベルの中からさらにラベルを選択するため、学習に用いるラベル数が減り、計算量の削減も期待できる。

3.2 類似性が低いラベルの選択法

ラベルの選択法は、ラベルのクラスタリングとラベルの選択の 2 つのステップから構成される。

3.2.1 ラベルのクラスタリング

提案法では、正解ラベルが持つトピックと共起しにくいトピックを見つけることが必要となる。そこで、本研究の特徴として、学習データセットに含まれるすべての異なるラベルを、各ラベルが有していないトピックの類似性でクラスタリングする。これにより、共起しにくいトピック集合でクラスタを作ることができるため、共起しにくいトピックを見つけるという目的に合致したクラスタリングを行うことができる。そのようなクラスタリングのために、ラベルの類似度には式 (10) を用いる。ラベルベクトル \vec{L} の各成分はラベル L 内でのトピックの有無を表わしており、本研究では各ラベルが持っていないトピックの類似度を測りたいため、トピック有を 0、無を 1 とする。例えば、「スポーツ」、「政治」、「社会」、「芸能」などのトピックがあり、 $L = \{\text{政治, 社会}\}$ の場合、 $\vec{L} = (1, 0, 0, 1)$ となる。

$$\text{sim}(L, L') = \frac{\vec{L} \cdot \vec{L}'}{|\vec{L}| |\vec{L}'|}. \quad (10)$$

クラスタリング手法には、式 (10) の類似度を用いた k -means 手法を使用する。本研究で用いる、クラスタ数を表わす k の値の設定方法について説明する。 k の値を適切に設定することで、共起しにくいトピックをうまく選択することができ、分類精度が向上する。適切な k の値は、データセット内に存在するラベル数とトピック数に影響を受けると考えられる。すなわち、ラベル数が多ければ k の値は大きくする必要があり、またトピック数が多ければ k の値は小さくする必要があるので、 k の値は式 (11) より設定する。

$$k = \frac{N_L}{N_T} \times \beta. \quad (11)$$

ただし、 N_L はデータセット内のラベル数、 N_T はトピック数、 β はパラメータである。

3.2.2 ラベルの選択

正解ラベルとの類似性が低い隣接ラベルほど誤分類に影響を与えるため、そのようなラベルを選択することを考える。3.2.1 節でクラスタリングした結果、正解ラベルが所属するクラスタ内のすべてのラベルが持っていないトピックは、正解ラベルが持つトピックと共起しにくいトピックである。そのため、隣接ラベルの中から、その共起しにくいトピックを持つラベルを選択し、学習に用いる。提案手法で選択されるラベル集合 B_i は式 (13) で表わされる。 $L_k (k = 1, \dots, N_i)$ は i 番目の学習データの正解ラベルが所属するクラスタ内のラベルを表わしており、 N_i はクラスタ内のラベル数である。 \bar{L} はラベル L の補集合を表わしている。

$$K_i = \{\bar{L}_1 \cap \bar{L}_2 \cap \dots \cap \bar{L}_{N_i}\}, \quad (12)$$

$$B_i = \{\lambda \in \Lambda \mid |\lambda \cap K_i| = 1 \cap \lambda \in A_i\}. \quad (13)$$

また、提案手法で選択されたラベルのみを学習に用いる最適化問題は次のように表わされる。

$$\begin{aligned} & \max_{\{\alpha_i^\lambda: \lambda \in B_i, i=1, \dots, m\}} \sum_{i=1}^m \sum_{\lambda \in B_i} \alpha_i^\lambda - \frac{1}{2} \sum_{i,j=1}^m \sum_{\lambda \in B_i, \lambda' \in B_j} \alpha_i^\lambda \alpha_j^{\lambda'} \\ & \times (\mathbf{x}_i \cdot \mathbf{x}_j) \frac{D(L_i, L_j) - D(L_i, \lambda') - D(\lambda, L_j) + D(\lambda, \lambda')}{2\sqrt{(1 - D(L_i, \lambda))(1 - D(L_j, \lambda'))}}, \\ \text{s.t. } & 0 \leq \alpha_i^\lambda \leq C. \end{aligned} \quad (14)$$

提案手法の更新 1 回あたりの計算量は $O(M_B^2)$ となる。ただし、 $M_B = \sum_{i=1}^m |B_i|$ である。 $|B_i| < |A_i|$ より $M_B < M_A$ となり、計算量が削減されている。

4 実験

提案手法の有効性を検証するため、マルチトピック文書分類で一般的に使用されている実験データセットである yahoo data を用いて、従来手法との予測精度と計算量の比較を行った。

4.1 実験条件

本実験では、Web ページのデータセットである yahoo data (詳しくは [6] を参照) を使用した。各 Web ページは 1 つ以上のトピックをもっており、マルチトピックとなっている。yahoo data は 11 個の異なるデータセットから構成されており、各データセットは表 1 の通りとなる。

各パラメータの設定は、事前に行った予備実験の結果より、 $C = 10$ 、 $\beta = 3$ とした。従来手法は学習データ数 1000 件で予測精度が安定したため、本実験の学習データ数を 1000 件とし、テストデータ数は 3000 件とした。従来手法の MML と提案手法を実装し、予測精度と計算量

表 1: データセット概要

データセット (略称)	文書数	トピック数
Arts&Humanities(Ar)	7484	26
Business&Economy(Bu)	11214	30
Computers&Internet(Co)	12446	33
Education(Ed)	12033	33
Entertainment(En)	12730	21
Health(He)	9206	32
Recreation(Rc)	12828	22
Reference(Rf)	8028	33
Science(Si)	6428	40
Social Science(SS)	12113	39
Society&Culture(SC)	14515	27

の比較を行った。予測精度の評価は、式 (15) で定義される平均ダイス係数 \bar{D} を使用した。

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n \frac{2|L_i^{true} \cap L_i^{pred}|}{|L_i^{true}| + |L_i^{pred}|}. \quad (15)$$

n はテストデータ数、 L_i^{true} は正解ラベル、 L_i^{pred} は予測ラベルを表わす。計算量は、各データセットにおいて M_A^2 と M_B^2 により、更新 1 回あたりを比較した。

4.2 実験結果と考察

図 2 に従来手法と提案手法の予測精度の比較を、図 3 に計算量の比較を示す。図 2 より、11 データセットの内 10 データセットで提案手法が従来手法より高い予測精度となっており、提案手法の有効性が示されている。提案手法では、誤分類に強い影響を与えるラベルを適切に選択できたため、予測精度が上がったと考えられる。また、クラスタリング後のクラスタ間の類似度が低いほど、予測精度が向上すると考えられるため、予測精度がより向上している Rf や Si はクラスタ間の類似度が低いと言える。Rc については、予測精度が下がってしまっているが、これはクラスタ数 k の値が小さすぎ、共起しにくいトピックが選択されなかったことが原因であることが追加実験で確認されている。クラスタ数 k の値を大きくすることで、予測精度は従来手法を上回る。よって、式 (11) の適切な β や k の設定方法など、クラスタリングのチューニング法を見直すことが課題として挙げられる。

他方、計算量の比較を示した図 3 より、更新 1 回あたりの計算量はすべてのデータセットにおいて提案手法が大きく削減されていることがわかる。また、 $M_B < M_A$ より提案手法の求めるパラメータ α_i^λ の数は少なくなっている。そのため、更新回数も提案手法が小さくなる場合が多く、全更新における計算量は更新 1 回あたりよりさらに大きく削減されていると言える。

以上のことから、提案手法は、予測精度の向上と計算量の削減という目的を達成することができ、その有効性を示すことができた。

5 まとめと今後の課題

本研究では MML の精度向上のため、隣接ラベルの中でも、特に誤分類に影響を与えるラベルに注目した。提案手法では、正解ラベルと類似性の低い隣接ラベルが誤分類に強い影響を与えると考え、そのラベルのみを選択して学習に用いることにより、予測精度の向上と計算量の削減を行った。類似性の低い隣接ラベルの選択法は、学習データセットに含まれるラベルをクラスタリングした後、正解ラベルが所属するクラスタ内でどのラベルも持って

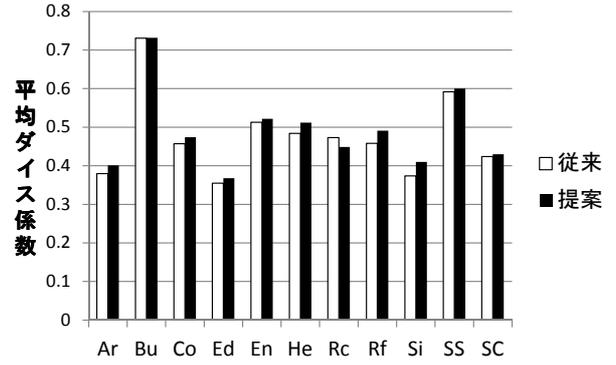


図 2: 予測精度の比較

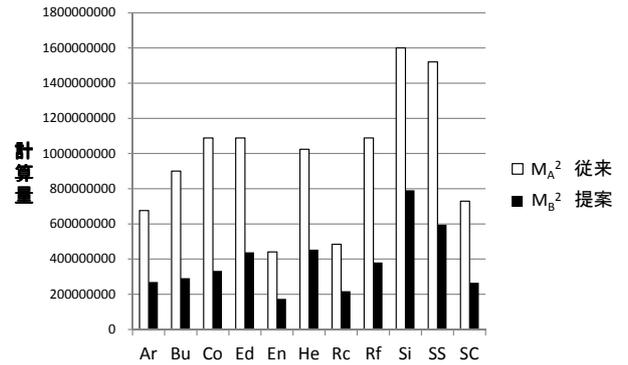


図 3: 計算量の比較

いないトピックを持つ隣接ラベルを選択し、そのラベルのみを学習に用いる。検証実験の結果、従来の MML と比較して、提案手法は予測精度の向上と計算量の削減を達成することができた。

今後の課題としては、検証実験より、11 個のデータセットの内 1 個だけ予測精度が下がってしまった問題に対する対処が挙げられる。クラスタ数 k の値が小さすぎたことが原因と考えられるため、クラスタリングにおいて k の値の設定方法などを見直すことで、問題の改善ができると考えられる。

参考文献

- [1] J. Thorsten, "Text categorization with support vector machines: Learning with many relevant features," Springer Berlin Heidelberg, 1998.
- [2] K. Miyahara, M. J. Pazzani, "Collaborative filtering with the simple Bayesian classifier," *Proc. the 6th Pacific Rim International Conference on Artificial Intelligence*, pp.679-689, 2000.
- [3] 賀沢秀人, 泉谷知範, 平博順, 前田英作, 磯崎秀樹, "最大マージン原理に基づく多重ラベリング学習," *電子情報通信学会論文誌 D*, 88(11), pp.2246-2259, 2005.
- [4] H. Kazawa, T. Izumitani, H. Taira, and E. Maeda, "Maximal Margin Labeling for Multi-Topic Text Categorization," *Advances in Neural Information Processing Systems 17*, pp.649-656, 2005.
- [5] C. Benhui, L. Ma, and J. Hu, "An Improved Multi-label Classification Method based on SVM with Delicate Decision Boundary," *International Journal of Innovative Computing, Information and Control* 6.4, pp.1605-1614, 2010.
- [6] N. Ueda and K. Saito, "Single-shot detection of multiple categories of text using parametric mixture models," *Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.626-631, 2002.
- [7] Lu Youwei, Shogo Okada, and Katsumi Nitta, "Semi-supervised Latent Dirichlet Allocation for Multi-label Text Classification," *Recent Trends in Applied Artificial Intelligence*, Springer Berlin Heidelberg, pp.351-360, 2013.