社団法人 情報処理学会　研究報告
IPSJ SIG Technical Report

2007－ＮＬ－179　(1)
2007－SLP－ 66　(1)
2007／5／24

トピック共起カーネルを利用した多重トピック文書分類

力徳正輝
(株) ジャストシステム イノベーティブテクノロジー研究開発部
masaki_rikitoku@justsystem.co.jp

マージン最大化法による多重トピック文書分類においてトピック共起カーネルをトピック素性空間に導入する．このカーネルはトピックの共起頻度から構成され，各トピック間の相関を取り込むことができる．またトピック共起カーネルは非対角項を持つカーネル行列として表現され，分類時には重みベクトルを明示的に構成できる．そのため分類時にもカーネル関数を必要とする手法に比べ高速な分類を実現できる．実験結果から提案カーネルは，線形カーネルに比べサイズの大きい多重トピックに対する平均トピック F1 値で優れていることがわかった．

キーワード:サポートベクターマシン，多重トピック文書分類，カーネル関数，Dice 係数，構造マッピング学習

Multi-topic Text Classification by Using Topic Co-occurrence Kernel

Masaki Rikitoku
Innovative Technology R&D Dept., Justsystems Corporation
masaki_rikitoku@justsystem.co.jp

In this paper, we propose the topic co-occurrence kernel for a multi-topic text classification based on the max-margin principle, which kernel is defined by a topic co-occurrence frequency in the topic feature space and can deal with correlation of the topics. The topic co-occurrence kernel is represented as a kernel matrix with non-diagonal matrix elements and we can explicitly construct the weight vector in a classification. It enables the faster classification compared with the methods requires a kernel in a classification.
Experiments show that the proposed kernel outperformed the linear kernel in the case that the number of multi-topics is greater than 1.

Keywords : SVM, multi-topic text classification, kernel, structured output, Dice coefficient

1　はじめに

文書分類において従来は，1 文書に 1 分類ラベルが付与されることが前提となっていた．これは排他的な分類が行われ，文書に 1 分類ラベルが付与されていた方が，その後のアプリケーションで利用しやすいためだと考えられる．しかしこの排他的な分類を実現するために分類器の性能向上はもちろん，厳格に構成された分類基準，分類ラベル定義が必要であり，この分類基準の構築とメンテナンスには多大のコストがかかるのが普通であった．

一方現在は，タグを使って文書を含む様々な情報を整理するサービスが広まってきている．その理由の 1 つは，上記の様な厳格な分類システムや複雑な分類階層構造を使用せずにフラットな分類タグを複数付与させるシンプルな仕様であると考えられる．しかし，付与された多重分類タグを使って情報検索を効率的に行うには付与された多重分類タグを精度良く付与，判定する必要がある．こういった背景から，文書分類において 1 文書に複数の分類ラベルを付与，認定する多重トピック文書分類の重要性は高まってきている．

賀沢らは [2],[7] において，多重トピック文書分類にマージン最大化法による分類手法 (MML) を提案した．彼らは多重トピックのベクトル間の類似度に相当するカーネル関数に線形カーネルとトピック F1 値に基づく非線型カーネルを使用し，一対他方式の SVM を含む他の多重トピック文書分類器との精度比較実験を行った．実験の結果，MML は複数の精度指標において他の手法

より優れていることが明らかになった．

MML は精度性能に優れた多重トピック文書分類器ではあるがいくつか改善するべき点がある．1 つは多重トピック分類の本来の目的である出力トピック数が大きい場合の精度性能である．彼らの報告では出力トピック数が 4 以上の場合においては上田らによって提案された多重トピック分類アルゴリズム PMM[6] と同等以下の性能を示すと報告している．さらに，トピック F1 値に基づく非線型カーネル関数を使用した場合，分類時にもカーネル関数を使って分類する必要があり，その分類処理速度の遅さから現実には実用が困難な点である．

本研究では MML の上記の 2 つの点を改善するためにマージン最大化多重トピック文書分類器に訓練データ中の各トピックの共起頻度に基づいて Dice 係数から定義されるトピック共起カーネルを適用することを提案する．トピック共起カーネルは非対角項を持つ線形カーネル関数として定義され，トピック間の相関を Dice 係数で採り入れることで上記の問題点を改善することができる．

本論文は次のように構成される．2 章で，一般的なマージン最大化による多重トピック文書分類と賀沢らの MML の説明を行う．3 章で本研究で提案するトピック共起カーネルを定義し，4 章で線形カーネルとトピック共起カーネルとの比較実験の結果を記述する．5 章で本論文のまとめを行う．

2 マージン最大化法による多重トピック文書分類

多重トピック文書分類とは，1 文書に複数の分類クラス，トピックを付与するタスクである．分類トピックの集合を $C = \{c_i \mid i = 1, \ldots, c\}$ とすると，付与される多重トピックは次の様なトピック素性空間のバイナリベクトル $\mathbf{y} = (y_1, \ldots, y_c)$ として表現できる．

$$y_m = (\mathbf{y})_m = \begin{cases} 1 & m \text{ 番目のトピックが付与される} \\ 0 & m \text{ 番目のトピックが付与されない} \end{cases}$$

例として，5 つの分類トピック集合 $C = \{$'スポーツ','国内','国際','政治','経済'$\}$ を持つ場合を考える．国際政治を扱っている文書に '国際' と '政治' トピックが付与される場合には 多重トピックベクトルは $\mathbf{y} = (0,0,1,1,0)$ と表現され，国際経済を扱っている文書に，'国際', '経済' が付与される場合には $\mathbf{y} = (0,0,1,0,1)$ と表現できる．

文書の bag of words 表現を $\mathbf{x} = (x_1, \ldots, x_N)$ とすると，多重トピック文書分類の訓練データセット D は $D = \{(\mathbf{y}_i, \mathbf{x}_i) \mid i = 1, \ldots, n\}$ と表現できる．以降，分類器の定式化においても，実験設定においても，文書の bag of words 表現は長さ 1 に正規化されているとする．

2.1 マージン最大化構造マッピングの学習

入力文に対して構文木を出力したり，文書に多重トピックを付与するといったような，入力データと構造を持っている出力データとの写像を推定するタスクは構造マッピング学習と呼ばれる．構造マッピング学習はいくつかの定式化があるが基本的には次の様に定式化できる [4].

入力対象と出力対象を 1 つの素性ベクトル $\Psi(\mathbf{y}, \mathbf{x})$ として考え, 次のスコア

$$\arg\max_{\mathbf{y}} \ score = \left\langle \mathbf{w}, \frac{\Psi(\mathbf{y}, \mathbf{x})}{|\Psi(\mathbf{y}, \mathbf{x})|} \right\rangle, \quad (1)$$

を最大にする $\mathbf{y}$ を出力対象する．ここで，$\langle \mathbf{a}, \mathbf{b} \rangle$ および $|\mathbf{a}|$ は考えている素性空間での内積とその内積によるベクトルの長さを表す.

重みベクトル $\mathbf{w}$ はマージン最大化に対応する次の最適化問題によって決められる．

$$\text{min.} \quad \frac{1}{2}\langle \mathbf{w}, \mathbf{w} \rangle + C \sum_{i}^{n} \xi_i \quad (2)$$

$$\text{s.t.} \quad \langle \mathbf{w}, \frac{\Psi(\mathbf{y}_i, \mathbf{x}_i) - \Psi(\mathbf{y}', \mathbf{x}_i)}{|\Psi(\mathbf{y}_i, \mathbf{x}_i) - \Psi(\mathbf{y}', \mathbf{x}_i)|} \rangle \geq 1 - \xi_i, \quad i = 1, \ldots, n, \quad \forall \mathbf{y}_i'. \quad (3)$$

ここで，ξ_i, C は例外的事例に対するスラック変数とそのコストパラメターである．

上記の最適化問題は，正解出力 $\mathbf{y}_i$ とそれ以外の可能な出力 (負例) $\mathbf{y}_i'$ とのマージンを最大化する重みベクトル $\mathbf{w}$ を決定する問題とみなせる．

一般に，Ψ は 入力データの素性空間と出力データの素性空間から作られるベクトル空間であるが，本研究ではそれぞれの素性空間のテンソル積として作られるベクトル空間とする．その場合, 内積がそれぞれの素性空間の内積の積と表現でき取扱いが簡易であるためである．

2.2 マージン最大化多重ラベリング学習

賀沢らは [2],[7] において，多重トピック文書分類を構造マッピング学習の枠組で定式化した．

彼らはトピック素性空間で線形カーネル，もしくは多重トピックの F1 値から決まるトピック F1 カーネル関数を使用し，語彙素性空間に対して線形カーネル関数を使用した．そして，Ψ としてトピック素性空間と語彙

素性空間のベクトルのテンソル積として表し多重トピック文書分類を定式化した．トピックF1カーネルは次式で定義される非線型カーネル関数である．

$$K(\mathbf{y},\mathbf{y}^{'}) = \frac{2\sum_{m=1}^{c} y_m \times y_m^{'}}{\sum_{m=1}^{c}(y_m + y_m^{'})} \tag{4}$$

学習において，各正解多重トピック $\mathbf{y}_i$ の負例の生成には，各トピック要素を反転，つまり，$y_m = 0$ ならば $y_m = 1$，逆ならば $y_m = 0$ として c 個の負例多重トピック $\mathbf{y}^{'}$ を生成する近似的な手法を採用し，現実的な学習処理時間を実現している．本研究でもこの負例作成法を採用する．

この負例作成手法によって，学習は次の双対問題表示の2次計画問題として定式化される．これはSVMの多値分類への自然な一般化となっている．

$$\begin{aligned} \text{min.} \quad & \frac{1}{2}\sum_{i,j=1}^{n}\sum_{l,m=1}^{c} \alpha_{il}\ \alpha_{jm}\hat{K}(\mathbf{y}_i - \mathbf{y}_l^{'}, \mathbf{y}_j - \mathbf{y}_m^{'})\langle \mathbf{x}_i, \mathbf{x}_j\rangle \\ & -\sum_{i}^{n}\sum_{l}^{c}\alpha_{il} \\ \text{s.t.} \quad & 0 \le \alpha_{il} \le C, \quad 1 \le i \le n,\ 1 \le l \le c \end{aligned} \tag{5}$$

ここで，$\hat{K}(,)$ は正規化カーネルを表し

$$\hat{K}(\mathbf{a},\mathbf{b}) = \frac{K(\mathbf{a},\mathbf{b})}{\sqrt{K(\mathbf{a},\mathbf{a})K(\mathbf{b},\mathbf{b})}}, \tag{6}$$

$\mathbf{y}_{il}^{'}$ は，i 番目の多重トピック $\mathbf{y}_i$ の m 番目のトピック要素を反転させ生成した負例多重トピックベクトルを表す．

(5) 式の最適化問題は2次計画問題であり，様々な効率的な解法アルゴリズムが存在する．しかし，多重トピック文書分類の場合，最適化する変数の個数は文書数 × トピックス数と大きくなり，通常はSVMの場合と同じく一般的な解法アルゴリズムの適用が難しくなる．そこでSVMのSequential Minimum Optimization(SMO)アルゴリズム [3] を拡張した (5) 式の解法アルゴリズムを本研究では開発適用したので次節で説明する．

2.3 拡張SMOアルゴリズム

SVMのSMOアルゴリズムは最急降下法を基本とする反復解法の1つである [3]．各反復で最も目的関数を下げる2変数を選択しその変数のみを逐次更新し最適解を得る．SVMの場合には等式制約式が1つあるために，自由に更新できる最小の変数である2変数を選択し逐次更新していくが，本研究の多重トピック分類器の場合には等式制約式がないため，1変数を選択，更新していく拡張SMOアルゴリズムを使用する．

拡張SMOアルゴリズム

Step 0. 誤差定数EPSに正の定数を設定，全ての $\alpha_{il} = 0$ に初期化する．
Step 1. violation 値 v_{il} が最大となる (i, l) を選択
Step 2. $v_{il} <$ EPS を満たしていれば終了，そうでなければ Step 3. へ
Step 3. α_{il} を (8) 式で更新
Step 4. α_{il} の更新に伴って全ての violation 値を更新して，Step 1. へ

violation 値 v_{il} は次で定義される．

$$v_{il} = -1 + \sum_{jm}\alpha_{jm}\hat{K}(\mathbf{y}_i - \mathbf{y}_{il}^{'}, \mathbf{y}_j - \mathbf{y}_{jm}^{'})\langle \mathbf{x}_i, \mathbf{x}_j\rangle. \tag{7}$$

これは各反復時での，i 番目の訓練事例ベクトルの分類誤りの大きさを示している．

1変数 α_{il} に対する更新式は (5) 式より

$$\alpha_{il}^{(new)} = -1 + \sum_{(j,m)\neq(i,l)}\alpha_{jm}\hat{K}(\mathbf{y}_i - \mathbf{y}_{il}^{'}, \mathbf{y}_j - \mathbf{y}_{jm}^{'})\langle \mathbf{x}_i, \mathbf{x}_j\rangle, \tag{8}$$

となる．

学習時は α の更新に伴って，全事例ベクトルに対する violation 値の更新処理が発生する．その際に (8) 式によってカーネル計算が行う必要あるため，そのままの実装では学習時間は膨大となる．SVMのSMOアルゴリズムの効率的な実装と同じくカーネル関数値のキャッシュと violation 値の差分更新を使用することで学習処理時間で短縮することができる．

3 トピック共起カーネル

多重トピック文書分類において，出力される多重トピックの中には相関を持つトピック対が存在し，それらは多重トピック分類において有効な分類素性になると考えられる．そこでトピック対の相関を取り入れたカーネルを構成する。

トピック対の相関を考慮するにはいくつかの可能性があるが，本研究では直感的に分かり易いカーネル行列の非対角項がDice係数で表されるトピック共起カーネルを提案する．

このトピック共起カーネルは共起の強さによって異なるトピック間の類似度を設定することができ，MMLの出力トピック数が大きい場合のトピックF1値を改善できることが期待できる．

3.1 Dice 係数によるトピック共起カーネル

多重トピックのなす素性空間において，非対角項を持つ行列による線形カーネル関数を

$$\langle \mathbf{y}, \mathbf{y}' \rangle_K = \mathbf{y}^\top K \mathbf{y}' = \sum_{i,j=1}^{c} K_{ij} y_i y_j', \tag{9}$$

とする．このカーネル関数は行列要素 K_{ij} を定めることで決まる．提案するトピック共起カーネルでは K_{ij} を Dice 係数を使って次の様に構成する．

$$K_{ij} = \frac{2 \times \sharp y_i \cap y_j}{\sharp y_i + \sharp y_j}, \quad i \neq j. \tag{10}$$

ここで，$\sharp y_i$ は，i 番目のトピックの訓練データ中の出現頻度であり，$\sharp y_i \cap y_j$ は，i 番目と j 番目のトピック対の共起頻度である．上記の定式化では，カーネル行列の対角項 K_{ii} は共起情報から決定することはできない．各トピックの出現頻度等から決定することも考えられるが，K がカーネル行列であるという要請から次の様に決めることができる．つまり K が正定値性を満たすために下三角行列 L を使って次の様に表されると仮定する．(付録 A 参照)

$$K = LL^\top \tag{11}$$

この様にカーネル行列を定式化すると、L と K の対角成分はよく知られている正定値対称行列の Choleskey 分解アルゴリズムから決めることができる．(付録 A 参照) ここで，L の対角成分は $L_{ii} = 1$ であるという仮定を置いた．これは任意のトピック共起行列が正定値性を満たし，K の非対角成分が小さい時には K は単位行列にほぼ同じであるという要請を考慮したものである．

3.2 トピック共起カーネルによる多重トピック分類

トピック共起カーネルは非対角項を持ったトピック素性空間での線形カーネル関数として定義したので，学習時には (5) 式にトピック共起カーネルをそのまま適用することができる．

分類時に実行されるスコア計算は，トピック素性空間の内積が (9) 式で定義されるため，

$$score = \frac{1}{|\mathbf{y}|_K} \langle \mathbf{w}, K\mathbf{y} \otimes \mathbf{x} \rangle, \tag{12}$$

となる．行列計算を含む (12) 式に基づいてスコア計算を分類対象毎に行うのは効率的ではない．そこで，トピック共起行列を繰り込んだ重みベクトルを

$$\mathbf{w}_K = \sum_{\alpha_{jm}>0} \alpha_{jm} \frac{K(\mathbf{y}_j - \mathbf{y}_{jm}')}{|\mathbf{y}_j - \mathbf{y}_{jm}'|_K} \otimes \mathbf{x}_j, \tag{13}$$

で構成すると，

$$score = \frac{1}{|\mathbf{y}|_K} \langle \mathbf{w}_K, \mathbf{y} \otimes \mathbf{x} \rangle, \tag{14}$$

となり，トピック間の相関を考慮しつつスコア計算は行列演算を含まない内積計算のみで実行できる．そのためカーネル関数を分類時にも利用する場合に比べて高速な分類処理を実現できる．

分類時には，各分類対象の長さによるスコアの正規化項を考慮したスコア最大化処理が多重トピック分類器には必要となる．それを含めた多重トピック分類アルゴリズムを下記に示す．

<u>**分類アルゴリズム**</u>

Step 0. 重みベクトル $\mathbf{w}_K$ を (13) 式で構築．
Step 1. 各分類対象文書 $\mathbf{x}$ に対して Step 2. 以下を順次実行．
Step 2. $m = 1, \ldots, c$ に対して，m 番目のトピックのみ $y_m = 1$ に設定した $\mathbf{y}_m$ に対して，(14) 式によりスコア計算しそれぞれのスコアを保持．
Step 3. トピックインデックス集合をスコアの降順にソート．
Step 4. 最大スコア $score_{max}$ を保持し，ソート済みトピックインデックス集合を保持．$k = 2$ に設定し Step 5. へ
Step 5. ソート済みトピックインデックス集合の上位 k 番目までのトピックの要素のみが 1 となる多重トピックベクトル $\mathbf{y}$ に対してスコア $score_k$ を計算し，Step 6. へ
Step 6. $score_k > score_{max}$ ならば，$score_{max} \leftarrow score_k$，$k \leftarrow k+1$ として，Step 5. へそうでなければ，ソート済みトピックインデクス集合の $k-1$ 番目までのトピックを多重トピックとして出力して，Step1. へ

多重トピックベクトルに対するスコアは線形カーネルを利用しているため，多重トピックの長さ $|\mathbf{y}|_K$ による正規化項を除いて，シングルトピックスコアの和になっていることに注意する．そのため多重トピックに対するスコア計算は，一般の線形分類器と同じシングルトピックに対する計算コストに次元の低いトピック素性空間での多重トピック $\mathbf{y}$ の正規化項計算コストとなり，比較的高速に分類を実行できる．

データサイズ	全体 LK	全体 TK	トピック数=1 LK	トピック数=1 TK	トピック数=2 LK	トピック数=2 TK	トピック数=3 LK	トピック数=3 TK	トピック数=4 LK	トピック数=4 TK	トピック数 ≥ 5 LK	トピック数 ≥ 5 TK
100	**49.14**	48.30	**54.45**	51.91	39.34	**41.51**	36.82	**39.30**	32.40	**34.92**	30.49	**33.15**
500	**58.69**	57.74	**63.53**	61.37	50.42	**51.55**	44.03	**45.64**	37.82	**39.74**	35.06	**36.90**
1000	**62.00**	60.19	**66.77**	63.89	**54.06**	54.05	46.11	**46.87**	39.45	**39.55**	36.47	**37.12**
2000	**64.25**	63.32	**68.92**	67.28	56.67	**56.96**	48.39	**49.11**	40.41	**41.17**	37.06	**38.45**

表 1: 多重トピックサイズ別平均トピック F1 値

4 実験

4.1 実験設定

トピック共起カーネルの効果を測定するために，マージン最大化法による多重トピック文書分類器において，線形カーネル (LK) とトピック共起カーネル (TK) との比較実験を行った．

実験には MML や PMM の実験で使用された yahoo data[5] を使用した．yahoo data は web page に対して 11 の大分類カテゴリと多重トピックが付与された多重トピック文書分類用のテストデータである．(詳細は，[6],[7] を参照) この yahoo data の 11 の大分類カテゴリ毎に，多重トピック文書分類を実行し精度評価を実行した．

それぞれの文書ベクトルは，tf $\times$ idf 値で素性化し，語彙素性空間の正規化カーネルを実現するために 2 乗長さで正規化し使用した．

実験には，オリジナルの訓練データからランダムに 2000 データを抽出，テストデータから同じく 3000 データをランダムに抽出し，2000 訓練データ+3000 テストデータを 1 つの実験タスクとした 5 セットの実験タスクを作成し学習，分類を行い 5 セットの平均を取り精度測定を行った．

多重トピック分類器のコストパラメータ C は，$C = 0.1, 1.0, 10.0$ と設定した予備実験において最良であった $C = 10.0$ を全ての実験で設定した．

使用した精度指標は，平均トピック F1 値，平均検索 F1 値，完全一致率の 3 つである．それぞれは $\{\mathbf{y}_1^{true}, \ldots, \mathbf{y}_n^{true}\}$ をテストデータの正解多重トピック，$\{\mathbf{y}_1^{pred}, \ldots, \mathbf{y}_n^{pred}\}$ を出力された多重トピックとして

$$\text{平均トピック F1} = \frac{1}{n}\sum_{i=1}^{n} \frac{2\sum_{j=1}^{c}(\mathbf{y}_i)_j^{true}(\mathbf{y}_i)_j^{pred}}{\sum_{j=1}^{c}((\mathbf{y}_i)_j^{true} + (\mathbf{y}_i)_j^{pred})},$$

$$\text{平均検索 F1} = \frac{1}{c}\sum_{j=1}^{l} \frac{2\sum_{i=1}^{n}(\mathbf{y}_i)_j^{true}(\mathbf{y}_i)_j^{pred}}{\sum_{i=1}^{n}((\mathbf{y}_i)_j^{true} + (\mathbf{y}_i)_j^{pred})},$$

$$\text{完全一致率} = \frac{1}{n}\sum_{i=1}^{n} I[\mathbf{y}_i^{true} = \mathbf{y}_i^{pred}], \qquad (15)$$

で定義される．ここで $I[true] = 1, I[false] = 0$ である．

4.2 実験結果

訓練データ数を 100，500，1000，2000 と変化させて平均トピック F1 値を測定し大分類カテゴリのマクロ平均を取った結果を表 1 に示す．

表 1 から，出力多重トピック数が 2 以上の場合では，ほぼ全ての場合で TK の平均トピック F1 値は LK を上回っている．特に訓練データサイズが小さい場合にはその差は顕著である．これは，トピック間の共起情報を多重トピック分類に有効な素性として利用できているためと考えられる．一方で，テストデータの大半を占める多重トピック数が 1 の場合では，TK の平均トピック F1 値は LK の結果を下回っており，そのため全体では TK は LK を下回る結果となっている

データサイズ	平均検索 F1 値 LK	平均検索 F1 値 TK	完全一致率 LK	完全一致率 TK
100	15.72	15.29	31.28	27.53
500	27.24	25.07	39.01	35.37
1000	32.06	30.16	42.21	38.60
2000	37.36	36.47	44.65	42.61

表 2: 平均検索 F1 値，完全一致率

表 2 に平均検索 F1 値，完全一致率の測定結果も示す．各精度は大分類カテゴリ毎の測定値の平均を取っている．平均検索 F1 値と完全一致率においては，全てのデータサイズで TK は LK を下回っている．表 1, 2 からトピック共起カーネルの使用は 1 文書に対する多重トピックの出力数を増加させトピックの再現率を向上させて平均トピック F1 値を向上させるが，適合率や完全一致率を向上させる効果は少ないことを示している．

5 まとめ

多重トピック分類器にトピック対の共起頻度から定義されるトピック共起カーネルを提案した．提案したトピック共起カーネルを使用した場合，出力多重トピック数が大きい場合の平均トピック F1 値を線形カーネルに比べて向上させることができた．また，提案したトピック共起カーネルは非対角項を持つカーネル行列として定義されているため分類時に重みベクトルを陽に構成することができ高速な分類を実行することが可能である．

トピック間の相関を多重トピック文書分類に採り入れる手法として提案したトピック共起カーネルは，いくつかの仮定やヒューリスティックによって定義されておりその理論的背景は明確ではない．今後は Fisher カーネル [1] の様にトピックの確率分布などに基づきトピック間の相関を理論的に自然に取り入れた多重トピック分類器とそのカーネルの研究を進めたいと考えている．

A Cholesky 分解によるトピック共起カーネルの構成

トピック共起行列 K が正定値条件を満たすよう次式の様に表されると仮定する．

$$K = LL^{\top}. \tag{16}$$

ここで 下三角行列 L は

$$L = (L)_{ij} = \begin{cases} 0 & \text{if } i < j \\ 1 & \text{if } i = j \\ l_{ij} & \text{otherwise} \end{cases}, \tag{17}$$

で定義される．

上式より

$$K_{ij} = \sum_{k=1}^{\min(i,j)} L_{ik}L_{jk}, \tag{18}$$

となる．

この仮定のもとで K の対角成分 K_{ii} と L の非対角成分 L_{ij} を Cholesky 分解アルゴリズムにより求めることができる．説明のために，逐次アルゴリズムを実行してみる．

まず，L_{i1} を求める．

$$K_{i1} = L_{i1}L_{11} \tag{19}$$
$$L_{i1} = K_{i1} \tag{20}$$
$$K_{11} = L_{11}L_{11} = 1 \tag{21}$$

となることがわかる．

次に L_{i2} は次の様に計算される．

$$K_{i2} = L_{i1}L_{21} + L_{i2}$$
$$L_{i2} = K_{i2} - L_{i1}L_{21} \tag{22}$$
$$K_{22} = (L_{21})^2 + 1 \tag{23}$$

上記を手順を順次実行していくことで，

$$L_{ij} = K_{ij} - \sum_{k=1}^{j-1} L_{ik}L_{jk}, \quad i > j \tag{24}$$

となり，L_{ij} を全て決定することができる．さらに K の対角成分は

$$K_{ii} = \sum_{k=1}^{i} L_{ik}^2 \tag{25}$$

と決定することができる．

参考文献

[1] T. Jaakkola, M. Diekhans, and D. Haussler. Using the Fisher kernel method to detect remote protein homologies. *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*, pages 149–158, 1999.

[2] H. Kazawa, Izumitani, H. Taira, and E. Maeda. Maximal Margin Labeling for Multi-Topic Text Categorization. *Advances in Neural Information Processing Systems 17*, pages 649–656, 2005.

[3] J.C. Platt. Fast training of support vector machines using sequential minimal optimization. *Advances in kernel methods: support vector learning table of contents*, pages 185–208, 1999.

[4] J. Rousu, C. Saunders, S. Szedmak, and J. Shawe-Taylor. On Maximum Margin Hierarchical Multi-label Classification. *NIPS Workshop on Learning With Structured Outputs*, 2004.

[5] N. Ueda. yahoo data file. `http://www.kecl.ntt.co.jp/as/members/ueda/publication-j.html`.

[6] N. Ueda and K. Saito. Single-shot detection of multiple categories of text using parametric mixture models. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 626–631, 2002.

[7] 平博順 前田英作 磯崎秀樹 賀沢秀人, 泉谷知範. 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-II*, pages 2246–2259, 2005.