カーネルリンク解析におけるパラメタ依存性と近似計算について

On Parameter Sensitivity Analysis and Approximate Computation for Kernel-based Link Analysis

新保 仁* 伊藤 敬彦† 松本 裕治‡

Masashi SHIMBO Takahiko ITO Yuji MATSUMOTO

本論文では, von Neumann カーネルをリンク解析に応用する際に問題となる, 二つの実際的な問題についてとりあげる. (1) リンク解析尺度として von Neumann カーネルの出力するランキングは, パラメタ依存性が高いことを指摘し, ランキング変動が起こる地点を効率的に推定する手法について論じる. (2) カーネル行列計算には, 逆行列を求める必要があるため, これを近似する複数の手法について議論する.

We discuss two issues that are crucial for practical application of von Neumann kernels to link analysis. (1) We demonstrate the parameter sensitivity of von Neumann kernels and propose a method for estimating parameter regions in which these kernels suddenly change their characteristics from a relatedness measure to an importance measure. (2) We also discuss approximate computation of kernel matrices to reduce time and space requirements, by avoiding expensive matrix inversion.

1. まえがき

PageRank [1] や HITS [6] に代表される Web 文書の重要度算出法は, 文書内容ではなく文書間のリンク情報を手がかりとして用いるため, **リンク解析尺度**と呼ばれる.

一方, 科学文献間の引用情報を活用して文献や学術論文誌, 著者間の関係を測定する手法は, 古くから計量書誌学分野において研究されており, 中でも, **共引用** (co-citation) や**書誌結合** (bibliographic coupling) といった文書間 '関連度' 指標は, 現在でも広く用いられている. 文献間の引用もまたリンクに他ならず, したがってこれらの関連度指標もリンク解析尺度とみなせる.

'重要度' および '関連度' 尺度は, 過去個別に議論されていたが, 我々は, グラフ上で定義されたカーネルを引用解析に適用することで, これらに対して統一的な解釈を与えた [4]. 具体的には, Kandola らの von Neumann カーネル [5] のパラメタ両端点にあたる二つのカーネルは, 各々, 共引用関連度と, HITS 権威度 (重要度) と一致し, したがって二つの端点間に存在 する任意のカーネルは, 関連度と重要度間の中間的なリンク解析尺度とみなすことができる. このことは, 単一のパラメタによって, 重要度あるいは関連度への偏りが調整できることを意味し, 画一的な重要度, 関連度のみを与える既存の方法に対して, より柔軟な運用や応用が可能になると期待できる.

ただし, von Neumann カーネルの実問題への応用を考えた場合, カーネルが (1) パラメタに非常に敏感である点, および (2) カーネル計算に伴う労力, が大きな問題となる. 本論文では, これらの点に関して, 簡単な解決法を提案し, 実験結果を報告する.

2. 予備知識

2.1 HITS 重要度

Kleinberg の HITS [6] は, PageRank と並んで, もっとも著名なリンク解析尺度のひとつである.

A を引用グラフの隣接行列とすると, **HITS アルゴリズム**は以下の再帰式を $n = 0, 1, 2, \ldots$ について計算する.

$$\mathbf{a}_{n+1} = \frac{A^{\mathrm{T}}\mathbf{h}_n}{|A^{\mathrm{T}}\mathbf{h}_n|}, \quad \mathbf{h}_{n+1} = \frac{A\mathbf{a}_{n+1}}{|A\mathbf{a}_{n+1}|}. \tag{1}$$

ただし, $\mathbf{a}_0$, $\mathbf{h}_0$ は全ての要素が 1 のベクトル. すると, 文書 i の**権威度**は**権威度ベクトル**$\lim_{n\to\infty} a_n$ の i-要素で与えられ, **ハブ度**は**ハブ度ベクトル** $\lim_{n\to\infty} h_n$ の i-要素で与えられる.

Kleinberg は, HITS の権威度ベクトルとハブ度ベクトルがそれぞれ, **共引用行列** $A^{\mathrm{T}}A$ と**書誌結合行列** AA^{T} の最大固有値に対応する固有ベクトルに一致することを示した.

2.2 von Neumann カーネル

Kandola らの von Neumann カーネル [5] は, B を対称行列として, 以下のように定義される.

$$N_\gamma(B) = B \sum_{n=0}^{\infty} \gamma^n B^n, \tag{2}$$

ここで $\gamma \geq 0$ は拡散係数と呼ばれるパラメタである. 元来, von Neumann カーネルは単語間あるいは文書間の関連度を求めることを目的としていたため, 単語-文書行列 X から得られる $B = X^{\mathrm{T}}X$, $B = XX^{\mathrm{T}}$ を用いたが, X の代わりに引用グラフの隣接行列 A を用いることで, リンク解析に用いることができる. この場合, B として, 共引用行列 $A^{\mathrm{T}}A$ または書誌結合行列 AA^{T} のいずれを用いるかに応じて 2 種類の von Neumann カーネルが得られる.

$$\hat{K}_\gamma = N_\gamma(A^{\mathrm{T}}A) \quad \hat{M}_\gamma = N_\gamma(AA^{\mathrm{T}}).$$

以下, 紙面の都合上, $\hat{K}_\gamma = N_\gamma(A^{\mathrm{T}}A)$ に話を限定する. 級数 $\hat{K}_\gamma$ は, $A^{\mathrm{T}}A$ のスペクトル半径 (最大固有値) を λ として, $\gamma < 1/\lambda$ のとき収束し, その解は

$$\hat{K}_\gamma = A^{\mathrm{T}}A(I - \gamma A^{\mathrm{T}}A)^{-1} \tag{3}$$

となる. このカーネルに関して, 以下の結果が得られている [4].

定理 1 共引用グラフが連結な場合, von Neumann カーネル $N_\gamma(A^{\mathrm{T}}A)$ の極限 $\gamma \to 1/\lambda$ において, カーネル行列の各行の与えるランキングは, $A^{\mathrm{T}}A$ の最大固有ベクトルに基づくランキング (HITS 権威度ランキング) と一致する.

* 正会員. 奈良先端科学技術大学院大学 情報科学研究科. shimbo@is.naist.jp
† 奈良先端科学技術大学院大学 情報科学研究科. takahi-i@is.naist.jp
‡ 奈良先端科学技術大学院大学 情報科学研究科. matsu@is.naist.jp

逆に下限 $\gamma = 0$ においては, 明らかに $\hat{K}_\gamma = A^{\mathrm{T}}A$ となり, カーネルの与えるランキングは, 共引用関連度と一致する.

3. パラメタ依存性

前章のグラフカーネルをリンク解析に適用する場合, 適切な拡散パラメタ γ を目的に応じて選択する必要があり, 可能であれば自動調整できることが望ましいが, この点に関して White と Smyth [9] は, パラメタ選択は本質的に主観的な作業にならざるを得ないと述べている[1]. 最適なパラメタの値を画一的な基準で決定することは不可能, という主張であるが, 我々もまた, この主張は妥当と考える.

例として, von Neumann カーネルを応用した技術論文推薦システムについて考える. 少数の '根論文' (ユーザが興味を持つ論文) が入力として与えられた際, システムはユーザが興味を持ちそうな他の論文を推薦する. このような場合には, ユーザが持つ根論文が属する分野への理解に応じて, 拡散係数 (重要度へ偏り) を変化させるべきである. すなわち, その分野について知識のないユーザに対しては, その分野における重要度のもっとも高い論文を, 分野に関する十分な専門知識を持つユーザに対しては, 重要度の高い論文よりはむしろ, 指定された根論文と関連度が高い論文を提示することが望ましい.

したがって, "最適" なパラメタを選択するにはユーザモデリングが必要となるが, これ自体が困難な問題であるため, 妥協案として, システムは, 複数のパラメタ値におけるランキングリストを出力し, 最適なランキングリストの選択は, 外部に位置するユーザモデリングツールに委ねることが考えられる.

このアプローチでは, 提示するランキングをパラメタのどの位置でサンプリングするか, という点が問題になる. なぜなら, 5.1 節の実験で示されるように, カーネルリンク解析尺度の性質は拡散パラメタに対し非線型に変化し, あるパラメタ区間ではランキングはほとんど変化せず, ごく短い区間において急激にランキングが変動する, という性質を持つ. したがって, 効率的なサンプリングを行うためには, ランキングの変化するパラメタ区間を見積もる指標が必要となる.

この問題に対して, 我々はパラメタ γ に対するカーネル行列の偏微分を使用する. カーネル行列の偏微分は以下のように解析的に解くことができる.

$$\frac{\partial N_\gamma(B)}{\partial \gamma} = (N_\gamma(B))^2, \tag{4}$$

B は固定されているため, ある γ における von Neumann カーネル N_γ が与えられたとき, 行列 $N_{\gamma+\Delta\gamma}$ の一次近似 $\tilde{N}_{\gamma+\Delta\gamma}$ は

$$N_{\gamma+\Delta\gamma}(B) \simeq \tilde{N}_{\gamma+\Delta\gamma}(B) = N_\gamma(B) + \Delta\gamma \frac{\partial N_\gamma(B)}{\partial \gamma}, \tag{5}$$

となる. ここで $\partial N_\gamma(B)/\partial\gamma$ は式 (4) より与えられる. 上式を用いて, $\tilde{N}_{\gamma+\Delta\gamma}(B)$ から導出されるランキングを $N_\gamma(B)$ から得られるランキングと比較することで, 区間 $[\gamma, \gamma+\Delta\gamma]$ におけるランキング変動を見積もることができる.

推定のコストは 式 (4), および (5) における行列積と和のみである. 式 (5) は, 式 (4) を用いると $N_\gamma(B)$ のみから計算できることに注意されたい. したがって提案手法では, サンプリングするのに適切な γ を確定するまで $N_{\gamma+\Delta\gamma}(B)$ を再計算する必要がなく, 二分探索等と組み合わせることで, 効率的なサンプリングが実現できる. この手法の精度については 5.2 節で実データを用いて検証する.

4. 近似手法

式 (3) から, von Neumann カーネルは逆行列計算によって求められることがわかる. このため, グラフ中に存在する節点数 n として, 計算量はほぼ[2] $O(n^3)$ となり, 大規模なグラフに適用する場合には問題となる. カーネル行列の近似計算手法として 2 種類の手法が考えられる.

4.1 主固有値による方法

まず, 主成分分析で行われているように, 行列の固有値上位 m 件によって近似する手法が考えられる. すなわち, 共引用行列 $A^{\mathrm{T}}A$ の固有値 $\lambda_1 \geq \ldots \geq \lambda_n$, そして対応する単位固有ベクトル $\mathbf{p}_1, \ldots, \mathbf{p}_n$ とすると, $P = [\mathbf{p}_1 \ \ldots \ \mathbf{p}_n]$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ として $A^{\mathrm{T}}A = P\Lambda P^{\mathrm{T}}$ と書け[3], したがって von Neumann カーネルは

$$N_\gamma(A^{\mathrm{T}}A) = PN_\gamma(\Lambda)P^{\mathrm{T}}$$

と書けるが, これを固有値対角行列 Λ の上位 $m(\leq n)$ 固有値のみに制限した行列 $\Lambda_m = \mathrm{diag}(\lambda_1, \ldots, \lambda_m, 0, \ldots, 0)$ と, P の対応する固有ベクトルのみを用いた $P_m = [\mathbf{p}_1 \ \ldots \ \mathbf{p}_m \ \mathbf{0} \ \ldots \ \mathbf{0}]$ で代用した,

$$\tilde{N}_\gamma(A^{\mathrm{T}}A) = P_m N_\gamma(\Lambda_m) P_m^{\mathrm{T}}$$

によって近似する.

後の実験で示すが, 拡散係数が上限に近い (すなわち, 重要度尺度に近い) 場合には, 重要度尺度が最大主固有ベクトルと一致することから, この手法はよい近似を与える. ただし, 拡散係数が小さい場合—すなわち, 関連度尺度としての性質が強い場合—には, 個々の節点に対して局所的な情報がより重要となるため, 小さな固有値も無視することができず, したがって, その近似精度は必ずしも良いとは限らない.

4.2 級数の k ステップ近似と空間使用量の削減

次に, von Neumann カーネル行列の無限級数表現 (式 (2)) の初項 k 項による近似が考えられる. 行列の近似理論 [3, §11.2] を用いると, この近似による誤差は $(|V|/k!)\,((\gamma\lambda)^{-1} - 1)^{-1/2}$ でおさえられることが示せる. ここで, λ は共引用行列のスペクトル半径 (最大固有値) である. もっとも, これは行列要素の誤差の上界にすぎず, カーネルから得られるランキングがどれだけ変化するかの見積りではない. このため, 5.3 章では, 近似手法によるランキングの変動予測精度を, 現実の科学文献引用グラフを対象に検証する.

また, 小規模な計算機を用いてカーネル行列を計算する際には, 計算時間もさることながら, 空間使用量も重大な問題となる.

[1] この White と Smyth の主張は, 彼らの提案するリンク解析尺度のパラメタに関して述べたものであるが, これはちょうど von Neumann カーネルの拡散係数に相当するパラメタであり, 後者に付いても当てはまる.

[2] 有名な Strassen の手法など, より計算量の小さな手法もあるが, 定数項が大きいため実用的ではない.

[3] $\mathrm{diag}(x_1, \ldots, x_n)$ で x_i を (i, i) -要素に持つ対角行列を表す.

表 1　von Neumann カーネルによるランキングと HITS 権威度ランキングの間の平均 K-min 距離
Table 1　Average K-min distance between rankings of von Neumann kernels and HITS

$\gamma\lambda$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99	0.999	0.9999	0.99999
Average K-min	87.3	87.2	87.1	86.9	86.3	85.6	84.3	81.7	72.0	26.4	5.5	1.1	0.0

引用グラフが疎であったとしても, 一般にカーネル行列は密となるため, グラフの規模が比較的小さな場合以外は, カーネル行列全体をメモリ上に確保することは困難である.

上記の k ステップ近似手法には, この空間使用量の削減が可能になる利点もある. カーネル行列全体をメモリ上に保存できない場合, もしくは, カーネル行列全体ではなく, 特定の節点 i に対する他の節点の重要度だけに興味がある場合には, カーネル行列の一部分のみを以下のようにして計算する. すなわち, $p = 1, \dots, k$ について, ベクトル $\left(\gamma A^{\mathrm{T}} A\right)^p \mathbf{u}_i$ を順次計算し, 合計する. これに $A^{\mathrm{T}} A$ を乗じたベクトル

$$A^{\mathrm{T}} A \sum_{p=1}^{k} \left(\gamma A^{\mathrm{T}} A\right)^p \mathbf{u}_i$$

によって, カーネル行列の i 列目に関する k ステップ近似を得る. ここで $\mathbf{u}_i$ は i-番目の要素だけが 1 という単位ベクトルである. この結果, 節点 i に関するランキング計算は, ベクトルの和とベクトルと行列の積となるが, これは HITS の繰り返し計算と本質的に同等であり, 十分実用に足ることがうかがえる.

5. 実験

自然言語処理に関する文献データベースより共引用行列を作成し実験を行った. これは [4] で用いた文献データベースから, 最大強連結成分を取り出したものであり, 節点 (この場合論文) 数計 2280 からなる. このように小規模なグラフを用いたのは, 解析解を計算し, 近似精度を見積もるためである. また, 強連結成分を取り出すことで, von Neumann カーネルの (パラメタ極限における) ランキングと HITS のランキングが完全に一致することを保証した.[4]

5.1 von Neumann カーネルのパラメタ依存性

von Neumann カーネルによって生成された各根論文に対する上位 10 件の論文リストと HITS による権威度ランキングとの間の相関を計測した. White と Smyth [9] に倣い, 相関の指標として, 最小化 Kendall (K-min) 距離 [2] を用いた. 表 1 は, 種々の γ に対する von Neumann カーネルのランキングと HITS ランキングとの K-min 距離を, 全 2280 根論文に対して平均をとったものである. K-min は距離尺度のため, 同一ランキングに対して 0 を, 不一致度の高いランキング間には大きな値を割り当てる. 今回, 上位 10 件を対象にしたため, 最大値は 100 となる.

表中, 拡散係数 γ は取りうる値の上限 $1/\lambda$ で正規化した値, つまり $\gamma\lambda$ として表示してある. ここで λ は共引用行列のスペクトル半径 (最大固有値). したがって $\gamma\lambda$ の取りうる値は, $0 \le \gamma\lambda < 1$ となる.

[4]HITS は, 共引用グラフが複数の強連結成分からなる場合, 1 個の強連結成分に含まれる節点のみに非 0 の権威度を割り当て, 他の成分の節点の権威度は全て 0 となる. von Neumann カーネルは全ての固有値を考慮して各接点毎にランキングを算出するため, そのような制限はない.

表 2　von Neumann カーネル N_γ と $N_{\gamma+\Delta\gamma}$ の K-min 距離 (厳密解) および, N_γ と式 (5) による一次近似 $\tilde{N}_{\gamma+\Delta\gamma}$ の K-min 距離 (推定値).
Table 2　K-min distance between von Neumann kernels N_γ and $N_{\gamma+\Delta\gamma}$ vs. K-min between N_γ and the approximated kernel $\tilde{N}_{\gamma+\Delta\gamma}$.

$\gamma\lambda$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
厳密解	0.27	0.20	0.30	0.47	0.88	1.32	2.32	7.11	70.26
推定値	0.07	0.13	0.22	0.34	0.65	0.99	1.56	3.31	9.74

(a) $\Delta\gamma = 0.099\lambda^{-1}$

$\gamma\lambda$	0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
厳密解	1.38	1.49	1.77	2.21	2.81	3.38	4.51	6.45	9.93	26.27
推定値	1.23	1.32	1.61	1.89	2.33	3.14	3.42	4.85	5.49	7.73

(b) $\Delta\gamma = 0.009\lambda^{-1}$

拡散係数 γ が増加するにしたがって, HITS ランキングとの差は減少し, $\gamma\lambda = 0.99999$ では完全に一致する. もっとも, 拡散係数パラメタの変化に対する, その変化量は決して線形ではないことに注意されたい. $0.1 \le \gamma\lambda \le 0.9$ ではそれほど変化は大きくないが, $\gamma\lambda > 0.9$ において K-min 距離は急激に減少し, von Neumann カーネルが与えるリンク解析尺度が重要度に近付いたことを示している.

5.2 パラメタ依存性推定

3 章で提案したパラメタ依存性の推定手法の妥当性を検証した.

表 2(a) の '厳密解' 行は, 拡散係数 γ における von Neumann カーネル N_γ による上位 10 件のランキングと, 拡散係数を $\Delta\gamma$ だけ増加させた場合のカーネル $N_{\gamma+\Delta\gamma}$ によるランキングとの K-min 距離を表示している. (2280 根論文に対する平均). ここで, $\Delta\gamma = 0.099\lambda^{-1}$ とした. この値は, 上位 10 件のランキングが区間 $[\gamma, \gamma + \Delta\gamma]$ でどれだけ変化するかを示している.

同様に, '推定値' 行は, N_γ と, 式 (5) による一次近似 $\tilde{N}_{\gamma+\Delta\gamma}$ との平均 K-min 距離を示す.

したがって, 同一 γ に対する, '厳密解', '推定値' との相関を見ることで, 近似精度を評価できる. この結果から, '厳密解' の値が大きく変化する区間では, '推定値' もまた相対的に大きくなっていることがわかる. 表 1 もあわせて参照のこと.

$\gamma\lambda = 0.9$ から 0.99 にかけての '厳密解', '推定値' の値の乖離が大きい (70.26 に対して推定値 9.74). この区間でのステップサイズが大きすぎたためであるが, 実際の運用の際には, 推定値しか入手できないため, この場合でも, この区間における推定値 9.74 のように他の区間 (最大でも 3.31) にくらべて相対的に推定値が大きいことが重要である. すなわち, 推定値が他区間と比較して大きい区間については, その区間でのステップサイズを減少させて, 再サンプリングを行う必要性があることを示している. 実際に $\Delta\gamma = 0.009\lambda^{-1}$ とステップサイズを小さくした結果を表 2(b) に掲げる. 乖離幅がより小さくなっているが, 必要に応じて, 推定値が最大の $[0.99, 0.999]$ の範囲ではさらにサンプリングを行えばよい.

表 3 von Neumann カーネルと, その k ステップ近似によるランキングの相違 (平均 K-min 距離).
Table 3 Average K-min distance between exact von Neumann kernels and its k-step approximation

	ステップ数 k					
$\gamma\lambda$	5	10	50	100	500	1000
0.1	0.0	0.0	0.0	0.0	0.0	0.0
0.9	10.0	6.0	0.1	0.0	0.0	0.0
0.99	50.7	42.5	14.9	6.5	0.1	0.0
0.999	75.1	66.5	31.6	21.2	4.8	1.6

表 4 カーネルと, その m 主固有値による近似を用いて得られたランキングの相違 (平均 K-min 距離).
Table 4 Average K-min distance between von Neumann kernels and its approximation by m-principal eigenvalues.

	固有値数 m				
$\gamma\lambda$	10	50	100	500	1000
0.1	63.3	39.5	28.4	17.4	0.7
0.9	52.9	36.0	24.4	7.1	0.7
0.99	23.6	16.7	11.7	9.7	6.1

5.3 近似計算

本節の実験では, べき級数表示の最初の k 項による近似計算の妥当性を検証する.

von Neumann カーネルのべき級数形式は, 式 (2), その解析解は (3) でそれぞれ与えられる. 初項 k 項による近似カーネルがそれぞれ生成するランキングどうしの K-min 距離を k を変化させて調べたのが表 3 である. 表中には表示されていないが, 標準偏差は $k = 1000$ において $\gamma\lambda = 0.999$ の von Neumann で 5.0, 他のカーネル, k に関しては 1 以下であった.

参考のため, 4.1 節の m 個の主固有値による近似結果についても, 表 4 に掲げる. $\gamma = 0.99$ の場合には, $m = 10$ 程度の固有値を用いても比較的良い近似が得られているが, $\gamma = 0.1$ と拡散係数が小さい場合, には, より多くの固有値が必要とされる.

6. 考察とむすび

カーネルリンク解析研究は, 緒に付いたばかりであり, 理論的な性質がようやく解明され始めたが, 実応用に用いる際に考慮すべき事項についての報告はほとんどなかった.

本論文では, これらの諸問題のうち, (1) パラメタ依存性への対処 (2) 効率的計算手法, について議論した. (1) に関しては, von Neumann カーネルがパラメタの変化に伴い大きく関連度尺度から重要度尺度へと性質が変化する点を推定する実用的な手法について論じた. (2) のカーネル計算に関しては, 現実の科学文献引用グラフにおいて k ステップ近似で良好な近似が可能であることを示し, また, その利点として, HITS 計算同様の手法によってカーネル行列の一部分のみを計算し, 空間使用量を削減することが可能であることを指摘した. 現実のリンク解析の応用においては, 特定の着目するグラフ節点に関する情報のみが得られれば十分である場合も多く, この空間使用量の削減は, そのような場合には極めて有用であると考える.

紙面が限られているため割愛したが, 提案したパラメタ依存性推定, 近似計算手法は, von Neumann カーネルのみならず, 正則化ラプラシアン [8], べき乗カーネル [5], 拡散カーネル [7] といった, その他のグラフカーネルについても適用可能である.

特に, べき乗カーネルと拡散カーネルには, 拡散パラメタに上界が存在せず, von Neumann カーネルのようにパラメタ上限近くで大きなランキングの変動 が起こる, といった性質が成り立たない. このため, これらのカーネルに対しては, 3 章のパラメタ依存性解析手法はより重要な意味を持つ.

さらに, ユーザの嗜好がベクトル (ランキング) という形で与えられたときに, パラメタ依存性解析手法を発展させ, パラメタの自動調整を行うことも考えられる. これについては今後の課題としたい.

[文献]

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual (web) search engine. *Computer Network and ISDN Systems*, 30(1–7):107–117, 1998.

[2] R. Fagin, R. Kumar, and D. Sivakumar. Comparing k lists. In *Proc. ACM-SIAM Symposium on Discrete Algorithms*, pp. 28–36, 2003.

[3] G. H. Golub and C. F. Van Loan. *Matrix Computation*. Johns Hopkins Univ. Press, Baltimore, 3rd ed., 1996.

[4] 伊藤, 新保, 工藤, 松本. カーネル法による計量書誌尺度の統一的解釈. 人工知能学会論文誌, 19(6):530–539, 2004.

[5] J. Kandola, J. Shawe-Taylor, and N. Cristianini. Learning semantic similarity. In *Advances in Neural Information Processing Systems 15*. MIT Press, Cambridge, MA, 2002.

[6] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46:604–632, 1999.

[7] R. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *Proc. 18th Intl. Conference on Machine Learning*, pp. 21–24, 2001.

[8] A. J. Smola and R. Kondor. Kernels and regularization of graphs. In *Proc. 16th Annual Conference on Computational Learning Theory*, pp. 144–158, 2003.

[9] S. White and P. Smyth. Algorithms for estimating relative importance in networks. In *Proc. 9th ACM SIGKDD Intl. Conference on Knowledge Discovery and Data Mining*, pp. 266–275, 2003.

新保 仁 Masashi SHIMBO

1994 年京都大学大学院大学院工学研究科修士課程電気工学第二専攻修了. 1997 年京都大学大学院工学研究科博士後期課程情報工学専攻指導認定退学. 博士 (工学). 現在奈良先端科学技術大学院大学学部情報科学学科助手.

伊藤 敬彦 Takahiko ITO

2004 年奈良先端科学技術大学院大学情報科学研究科情報処理学専攻修了. 現在同大学院大学博士後期課程に在学中. 学術振興会特別研究員 DC2.

松本 裕治 Yuji MATSUMOTO

1979 年京都大学大学院工学研究科修士課程情報工学専攻修了. 現在奈良先端科学技術大学院大学教授. 工学博士. 専門は自然言語処理.