

Markov Cluster Algorithmを用いた Web コミュニティ群の発見手法

加藤 一 民[†] 松尾 啓 志[†]

Web コミュニティとは、共通のトピックを扱う Web ページの集合を意味する。Kleinberg の HITS アルゴリズムは、Web ページ間のリンク関係を解析することにより、特定のトピックに関連する Web コミュニティを発見する。しかし、一般に広い意味を持っているトピックにおいては、トピックに関連するコミュニティは1つだけでなく、複数存在するため、単一のコミュニティを抽出する HITS アルゴリズムでは、ユーザが意図する Web コミュニティを抽出することができない場合がある。そこで本研究では、Markov Cluster Algorithm を用いて、ユーザから与えられたキーワード (トピック) から、それに関連する複数の Web コミュニティを発見する手法を提案する。

The Discovery Method of Multiple Web Communities with Markov Cluster Algorithm

KAZUTAMI KATO[†] and HIROSHI MATSUO[†]

A web community is a set of web pages created by individuals or associations with a common interest on a topic. Kleinberg's HITS algorithm find a web community on a query topic by link analysis. For multiple meanings of query terms, there might be multiple web communities related to the query topic. However, HITS algorithm can't always extract web community which users would expect or prefer, because it can only extract single community on a query topic. In this paper, we propose a method for discovering web communities on a query topic, using Markov Cluster Algorithm.

1. はじめに

近年、インターネットの発展により、Web ページの数は増加の一途を辿っており、今後もこの傾向は続くと考えられる。また、Web の拡大に伴って、ユーザが Web 上から必要な情報を見つけ出すことがますます困難になってきている。そこで、膨大な量の情報の中からユーザが必要な情報を効率良く獲得するための工夫として、意味的に関連した Web ページ群をひとまとまりのグループとして扱う方法がある。このような Web ページのグループ化に関する研究の一つとして、関心や興味を共有する人々によって作成された Web ページの集合 (Web コミュニティ) を発見する研究がある。Web コミュニティの概念は、実社会において同じ目的や関心を共有する人々が集まってコミュニティを形成しているように、Web 上においても関心や興

味を共有する人々によって作成された Web ページが集まってコミュニティを形成しているという考え方に基づいている。

Web ページ間のリンク構造を解析して Web コミュニティを発見する手法としては、Kleinberg の HITS(Hyperlink-Induced Topic Search) アルゴリズム¹⁾ や、Web ページの参照の共起性に基づいた村田の手法^{2),3)} などが提案されている。Kleinberg が提案した HITS アルゴリズムは、特定のトピックに関連する Web コミュニティを抽出する手法である。HITS アルゴリズムには、アンカーテキスト、リンクへの重み付けなどを用いた様々な改善手法^{4)~6)} がこれまでに提案されている。

一般に広い意味を持っているトピックは、複数の意味が異なるトピックと密接に関連している。例えば、トピック apple は、アメリカの Apple Computer 社やフルーツの apple など複数の意味が異なるトピックに関連している。即ち、一般に広い意味を持っているトピックにおいては、トピックに関連するコミュニティは1つだけでなく、複数存在することが想定される。

[†] 名古屋工業大学大学院情報工学専攻
Department of Computer Science and Engineering,
Graduate School of Engineering, Nagoya Institute of
Technology

しかし、従来の Web コミュニティ発見手法では、トピックに関連するコミュニティを1つだけしか抽出することができないため、一般に広い意味を持っているトピックにおいては、ユーザが期待する結果が得られない場合がある。

本論文では、HITS アルゴリズムと Markov Cluster Algorithm (MCL) を組み合わせることにより、ユーザから与えられたキーワード (トピック) から、それに関連する複数の Web コミュニティを発見する手法を提案する。本論文で提案する手法は、与えられたトピックに関連するコミュニティを1つだけ抽出する従来の手法に比べて、トピックに関連する複数のコミュニティを提供することができるという点で優れている。

以下、2章では HITS アルゴリズムとその問題点について述べる。3章では、Markov Cluster Algorithm について概説し、本研究で提案する Web コミュニティ群の発見手法について述べる。4章では実験により提案手法の性能評価を行う。5章でまとめと今後の課題について述べる。

2. HITS アルゴリズム

HITS アルゴリズムは、Web ページ間のリンク関係を解析することにより、特定のトピックに関連する Web コミュニティを発見する手法である。HITS アルゴリズムの特徴は、“authority” と “hub” という、Web ページの有用性を表す 2 つの評価尺度を導入している点である。“authority” は、特定のトピックに関する情報が豊富であることを表す尺度であり、“hub” は、authority としての価値が高いページへのリンクが豊富であることを表す尺度である。

各ページ p の authority と hub の評価値は、以下の 2 式を反復することにより計算される。

$$auth(p) = \sum_{q, q \rightarrow p} hub(q) \quad (1)$$

$$hub(p) = \sum_{q, p \rightarrow q} auth(q) \quad (2)$$

即ち、authority と hub の関係は、authority の評価値が高いページは、多数の hub の評価値が高いページからリンクを張られており、また hub の評価値が高いページは、多数の authority の評価値が高いページにリンクを張っているというように、再帰的に定義される。

本論文では、HITS アルゴリズムが出力する上位の authority、hub ページの集合を Web コミュニティであると定義する。このような authority と hub からなる Web コミュニティを抽出する HITS アルゴリズムの詳細を以下に示す。

1. あるキーワード σ を Yahoo⁸⁾ などの検索エンジンに入力し、そのキーワードを含む Web ページを t 件収集し、root 集合 (R_σ) とする。
2. root 集合に含まれる各ページ p に対して、 p にリンクを張っているページを最大 d 件収集し、backward 集合とする。
3. root 集合に含まれる各ページ p に対して、 p からリンクされているページをすべて収集し、forward 集合とする。
4. backward 集合と forward 集合を root 集合に追加して大きさ n の base 集合 (S_σ) を作成する。
5. base 集合内のすべてのページ間のリンク構造を解析してグラフを作成する。さらに、同じドメイン下にあるページ間のリンクをすべて削除し、異なるドメイン下にあるページ間のリンクだけを残したグラフ (G_σ) を作成する。
6. グラフ G_σ におけるページ間のリンク関係を示す $n \times n$ の隣接行列 A 、その転置行列 A^T を作成する。
7. $A^T A$ の最大固有値 μ_1 に対応する正規化された固有ベクトル x と、 AA^T の最大固有値 λ_1 に対応する正規化された固有ベクトル y を反復計算により求める。(固有ベクトル x, y の各要素の値が、各ページの authority, hub の値を示す)
8. 上位 M 個の authority と hub ページの集合を Web コミュニティとして抽出する。

2.1 HITS アルゴリズムの問題点

一般に広い意味を持っているトピックは、複数の意味が異なるトピックと密接に関係している。例えば、トピック apple は、アメリカの Apple Computer 社とフルーツの apple という 2 つの意味が異なるトピックに関係している。従って、一般に広い意味を持つキーワード σ を基にして作成された base 集合のグラフ G_σ には、複数のトピックに関連する Web ページ群がそれぞれ密なリンク構造を持っていることが考えられる。

図 1 は、森本らの手法⁹⁾により、トピック apple に対する base 集合のグラフ G_σ を可視化したものである。この手法では、リンクで接続されているノード (隣接ノード) は近くに配置し、リンクで接続されていないノード (非隣接ノード) は離して配置するという考え方に基いて、グラフを可視化している。同図より、トピック apple に対する base 集合のグラフは、複数の密なリンク構造を持つ Web ページ集合で構成されていることを確認することができる。

HITS アルゴリズムは、グラフ G_σ において、最も密なリンク構造を持つ Web ページ集合を authority と hub からなる Web コミュニティとして抽出する。

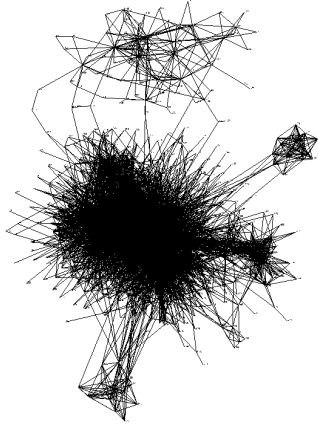


図1 トピック“apple”のWebグラフ
Fig. 1 Web graph of the topic “apple”

従って、グラフ G_σ が複数の密なリンク構造を持つ Web ページ集合で構成される場合、トピックに関連する Web コミュニティを1つだけしか抽出することができない。

一般に広い意味を持っているトピックにおいては、トピックに関連するコミュニティは1つだけではなく、複数存在するため、単一のコミュニティを抽出する HITS アルゴリズムでは、ユーザが満足する結果を提供することには限界があると考えられる。そこで本論文では、トピックに関連する複数の Web コミュニティを発見する手法を提案する。

3. 提案手法

HITS アルゴリズムでは、トピック σ を基にして作成された base 集合のグラフ G_σ において、最も密なリンク構造を持つ Web ページ集合をコミュニティとして抽出するため、トピックに関連するコミュニティを1つだけしか抽出することができない。一方、グラフ G_σ においては、同じトピックを扱っている Web ページ集合は、多数のハイパーリンクによって結合されているのに対して、異なるトピックを扱っている Web ページ集合の間には、少数のハイパーリンクしか張られていない。従って、クラスタ内部に多数のハイパーリンクを持ち、クラスタ間には少数のハイパーリンクしか持たないように、グラフ G_σ をクラスタリングすることで、意味が異なるトピックごとにクラスタを生成することができる。

提案手法では、Markov Cluster Algorithm(MCL)⁷⁾を用いてグラフ G_σ をクラスタリングし、複数のクラスタを生成する。そして、生成された各クラスタに対して HITS アルゴリズムを適用することによって、複数の Web コミュニティを抽出する。

3.1 Markov Cluster Algorithm

MCL は、グラフ上のランダムウォークに基づいたグラフクラスタリング手法であり、グラフ G に対する MCL は、以下の手順で行われる。

1. G における全てのノードに自己ループを追加する
2. $M = M_G$ (M_G は G に対するマルコフ行列)
3. Do
 - (a) $M = M^2$
 - (b) $M = \Gamma_r M$
 While ($M^2 \neq M$)
4. 行列 M に対応するグラフを H とする。 H の連結成分を1つのクラスタとして、クラスタ集合 $C = \{C_1, C_2, \dots, C_N\}$ (N :クラスタ数) を生成する

まず、グラフ G に対するマルコフ行列 M_G を作成する。ランダムウォークモデルで考えると、マルコフ行列 M_G はマルコフ連鎖の推移確率行列とみなすことができ、 $(M_G)_{pq}$ はノード q からノード p へ推移する確率を表す。また、マルコフ連鎖の k ステップ後の推移確率行列は、行列 M_G の k 乗 (M_G^k) で表すことができる。ここで、推移確率 $(M_G^k)_{pq}$ は、ノード p と q が存在する局所領域内のエッジの密度が高いほど大きくなる。MCL の基本的なアイデアは、expansion と inflation の2つの演算を交互に実行することで、この性質を強調することにある。行列 M に対する expansion 演算は、 $M = M^2$ で与えられる。また、行列 $M \in \mathbf{R}^{k \times l}$ に対する inflation 演算 Γ_r は、次式で定義される。

$$(\Gamma_r M)_{pq} = (M_{pq})^r / \sum_{i=1}^k (M_{iq})^r \quad (3)$$

r は inflation パラメータと呼ばれ、 $r > 1$ である。通常は $r = 2$ に設定される。

MCL は解の収束が保証されていないが、多くの実例(とりわけ無向グラフ)では、解が収束することが実験により示されている。図2に、MCLによるグラフクラスタリングの例を示す。

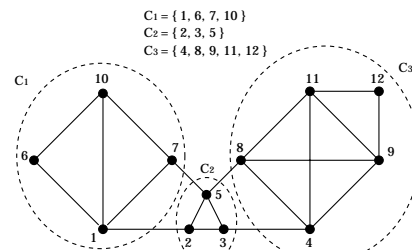


図2 MCLによるグラフクラスタリングの例
Fig. 2 Example of graph clustering with MCL

3.2 Web コミュニティ群発見アルゴリズム

Web コミュニティ群発見アルゴリズムは、キーワードを入力として、authority と hub からなる複数の Web コミュニティを出力する。以下に、アルゴリズムの詳細を示す。

1. キーワード σ を基にして base 集合のグラフ G_σ を作成する。(base 集合の作成手順は HITS アルゴリズムと同様)
2. グラフ G_σ を無向グラフ G'_σ に置き換える。MCL を用いてグラフ G'_σ をクラスタリングし、クラスタ集合 $C = \{C_1, C_2, \dots, C_N\}$ (N : クラスタ数) を生成する。
3. グラフ G_σ の頂点集合の部分集合 C_1, C_2, \dots, C_N により誘導された部分グラフを G_1, G_2, \dots, G_N とする。
4. 各 G_i について、HITS アルゴリズムを適用して、 G_i に含まれる Web ページの authority と hub の評価値を計算する。
5. 各 G_i について、上位 M 個の authority と hub ページの集合を Web コミュニティとして抽出する。

4. 実験

提案手法の有効性を確認するために、“apple”, “amazon”, “motor company” の3つのトピックに対して Web コミュニティ群を発見する実験を行い、HITS アルゴリズムと結果を比較した。これらのトピックは、いずれも広い意味を持っており、提案手法の有効性を効果的に示すことができると考えた。

4.1 base 集合のフィルタリング

HITS アルゴリズムおよび提案手法では、トピックによっては 10,000 ページ以上に及ぶ base 集合が作成される。HITS アルゴリズムと MCL アルゴリズムの計算量は、base 集合に含まれる Web ページ数 n の3乗に比例するため、大規模なグラフに対しては適しているとは言えない。そこで、HITS アルゴリズムおよび提案手法では、計算量を削減するため、野村らの手法⁶⁾を用いて base 集合を作成する段階で Web ページのフィルタリングを行った。

4.2 パラメータ設定

提案手法におけるパラメータ設定を表 1 に示す。base 集合のフィルタリング係数は、できるだけトピックに関係のあるページがフィルタリングされないようにするため $k = 2$ とした。また、HITS アルゴリズムにおいて、base 集合の作成に関するパラメータ設定は、提案手法と同じである。

表 1 提案手法のパラメータ

Table 1 Parameters for the proposed method

パラメータ	値
検索エンジン	米 Yahoo!
root 集合サイズ t	200
backward 集合サイズ d	50
base 集合のフィルタリング係数 k	2
inflation パラメータ r	1.4
コミュニティサイズ M	20

4.3 性能の比較と評価

HITS アルゴリズムと提案手法の実験結果を比較することにより、提案手法の性能を主観的に評価する。

[トピック apple の実験結果]

表 2 に、HITS アルゴリズムによる “apple” の authority 上位 5 件を示す (左の数値は、authority 値)。また、表 3 に、提案手法による “apple” の authority 上位 5 件を示す (クラスタサイズが大きいものから順に 1st, 2nd, 3rd,... と表記する)。

表 2 HITS アルゴリズムによる “apple” の実験結果

Table 2 An experimental result of “apple” by HITS algorithm

top 5 authorities	
.306	http://www.apple.com
.264	http://www.info.apple.com
.235	http://quicktime.apple.com
.217	http://kbase.info.apple.com
.209	http://kbase.info.apple.com/index.jsp

表 3 提案手法による “apple” の実験結果

Table 3 An experimental result of “apple” by proposed method

top 5 authorities (1st cluster)	
.306	http://www.apple.com
.264	http://www.info.apple.com
.236	http://quicktime.apple.com
.218	http://kbase.info.apple.com
.210	http://kbase.info.apple.com/index.jsp
top 5 authorities (2nd cluster)	
.484	http://www.bestapples.com
.456	http://www.nyapplescountry.com
.448	http://www.michiganapples.com
.398	http://www.usapple.org
.231	http://www.urbanext.uiuc.edu/apples
top 5 authorities (3rd cluster)	
.374	http://www.apple2.org
.367	http://www.a2central.com
.355	http://www.wbwip.com/a2web/a2webbring.html
.325	http://apple2history.org
.323	http://www.syndicomm.com/a2web/a2hmpgs.html

HITS アルゴリズムでは、第 1 位の authority は米

Apple Computer 社のホームページとなり、第 2 位以降も同社の関連ページが続いた。即ち、トピック “Apple Computer” に関するコミュニティが抽出された。

提案手法では、1 番目のクラスタにおいては、HITS アルゴリズムと同様に、米 Apple Computer 社の関連ページが上位 authority として抽出され、トピック “Apple Computer” に関するコミュニティとなった。さらに、2 番目のクラスタにおいては、第 1 位の authority はアメリカ・ワシントン州りんご協会のページ、第 2 位はアメリカ・ニューヨーク州りんご協会のページとなり、第 3 位以降もアメリカを中心としたりんごの関係機関となった。即ち、フルーツの “apple” に関するコミュニティが抽出された。また、3 番目のクラスタにおいては、米 Apple Computer 社が開発したパソコン appleII の関連ページが上位 authority として抽出され、トピック “appleII” に関するコミュニティとなった。

HITS アルゴリズムでは、トピック “Apple Computer” に関するコミュニティだけが抽出されたのに対して、提案手法では、“Apple Computer”，フルーツの “apple”，“appleII” という 3 つのトピックに関するコミュニティが抽出された。この結果から、提案手法において MCL によるグラフクラスタリングが有効に働いたと考えられる。

また、トピック “appleII” に関するコミュニティが抽出されたことは予想外の結果であった。提案手法では、ユーザから与えられたトピックを基にして、その近傍のトピックに関する Web ページ群をコミュニティとしてユーザに提示する。よって、本手法は、ユーザがあるトピックについて幅広く調べたい場合などに、有効な検索手段であると言える。

[トピック amazon の実験結果]

表 4 に、HITS アルゴリズムによる “amazon” の authority 上位 5 件を示す (左の数値は、authority 値)。また、表 5 に、提案手法による “amazon” の authority 上位 5 件を示す。

表 4 HITS アルゴリズムによる “amazon” の実験結果
Table 4 An experimental result of “amazon” by HITS algorithm

top 5 authorities	
.136	http://safari.oreilly.com
.135	http://www.xml.com
.135	http://webservices.xml.com
.135	http://digitalmedia.oreilly.com
.135	http://www.perl.com

HITS アルゴリズムでは、トピック “amazon” と

表 5 提案手法による “amazon” の実験結果

Table 5 An experimental result of “amazon” by proposed method

top 5 authorities (1st cluster)	
.505	http://www.amazon.com
.427	http://www.amazon.co.uk
.404	http://www.amazon.fr
.364	http://www.amazon.de
.338	http://www.amazon.co.jp
top 5 authorities (2nd cluster)	
.536	http://www.eduweb.com/amazon.html
.404	http://www.worldwildlife.org/amazon
.327	http://www.pbs.org/journeyintoamazonia
.266	http://www.extremescience.com/AmazonRiver.htm
.249	http://www.ran.org
top 5 authorities (3rd cluster)	
.136	http://safari.oreilly.com
.135	http://www.xml.com
.135	http://webservices.xml.com
.135	http://digitalmedia.oreilly.com
.135	http://mac.oreilly.com
top 5 authorities (6th cluster)	
.728	http://www.amazonation.com
.664	http://www.myrine.at/Amazons
.092	http://tx.essortment.com/amazonwarrior_ryci.htm
.092	http://www.ifi.uio.no/thomas/lists/amazon-connection.html
.061	http://www.politicalamazon.com

は全く関係のない、米オライリー社のページが上位 authority として抽出された。このように、HITS アルゴリズムの出力が本来のトピックとは全く関係のない Web コミュニティに収束してしまうという現象は、topic drift 問題⁴⁾と呼ばれており、従来から指摘されている。

提案手法では、1 番目のクラスタにおいては、電子商取引企業の米 Amazon.com のページが上位 authority として抽出され、“Amazon.com” に関するコミュニティとなった。第 1 位～第 5 位の authority は、同企業のアメリカ、イギリス、フランス、ドイツ、日本向けサイトである。2 番目のクラスタにおいては、南米エクアドルのアマゾン流域の関連ページが上位 authority として抽出され、南米の “amazon” に関するコミュニティとなった。3 番目のクラスタにおいては、HITS アルゴリズムと同様に、トピック “amazon” とは全く関係のない、米オライリー社のページが上位 authority として抽出された。さらに、4 番目のクラスタにおいても、トピック “amazon” とは全く関係のない、米 Jupitermedia 社のページが上位 authority として抽出された。5 番目のクラスタにおいては、1 番目のクラスタと同様に、Amazon.com の関連ページが上位 authority として抽出され、“Amazon.com” に関するコミュニティとなった。また、6 番目のクラスタにお

いては、ギリシャ神話に登場する女性だけの部族アマゾン (アマゾネス) の関連ページが上位 authority として抽出され、女性民族の “Amazon” に関するコミュニティとなった。

HITS アルゴリズムでは、topic drift 問題が発生したため、トピック “amazon” とは全く関係のない Web コミュニティが抽出されたのに対して、提案手法では、トピック “amazon” とは無関係のコミュニティが抽出される一方で、“Amazon.com”, 南米の “amazon”, 女性民族の “Amazon” という3つのトピックに関するコミュニティが抽出された。この結果から、提案手法は、base 集合に本来のトピックとは全く関係のない Web ページ群が含まれている場合でも、それらの Web ページ群をコミュニティの1つとして分離することができる。また、ユーザは提示された結果の中から自分に必要なコミュニティを選択するだけでいいため、比較的容易に検索を行うことができる。

[トピック motor company の実験結果]

表 6 に、HITS アルゴリズムによる “motor company” の authority 上位 5 件を示す (左の数値は、authority 値)。また、表 7 に、提案手法による “motor company” の authority 上位 5 件を示す。

表 6 HITS アルゴリズムによる “motor company” の実験結果
Table 6 An experimental result of “motor company” by HITS algorithm

top 5 authorities	
.320	http://www.honda.com
.298	http://www.toyota.com
.285	http://www.ford.com
.281	http://www.pontiac.com
.274	http://www.cadillac.com

HITS アルゴリズムでは、第 1 位の authority はアメリカ・ホンダのオフィシャルサイト、第 2 位の authority はアメリカ・トヨタのオフィシャルサイト、第 3 位以降もフォード、ポンティアック、キャデラックなどの自動車メーカーのオフィシャルサイトが続いた。即ち、トピック “motor car company” に関するコミュニティが抽出された。

提案手法では、1 番目のクラスタにおいては、HITS アルゴリズムと同様に、自動車メーカーのオフィシャルサイトが上位 authority として抽出され、“motor car company” に関するコミュニティとなった。2 番目のクラスタにおいては、ハーレー・ダビッドソンやビューエルなどのモーターサイクルメーカーのオフィシャルサイトが上位 authority として抽出され、“motorcycle company” に関するコミュニティとなった。また、3 番目のクラスタにおいては、Vintage、Chesil、

表 7 提案手法による “motor company” の実験結果
Table 7 An experimental result of “motor company” by proposed method

top 5 authorities (1st cluster)	
.319	http://www.honda.com
.299	http://www.toyota.com
.284	http://www.pontiac.com
.283	http://www.ford.com
.276	http://www.cadillac.com
top 5 authorities (2nd cluster)	
.452	http://www.harley-davidson.com
.432	http://www.buell.com
.326	http://www.confederate.com
.322	http://www.twineagle.com
.318	http://www.wildwestmc.com
top 5 authorities (3rd cluster)	
.742	http://www.vintagemotorcompany.co.uk
.420	http://www.leighton-cars.co.uk
.418	http://www.chesil.co.uk
.215	http://www.scampmotorcompany.co.uk
.195	http://www.jimini-cars.co.uk

Scamp などのイギリスを中心としたキットカーメーカーのオフィシャルサイトが上位 authority として抽出され、“kit car company” に関するコミュニティとなった。

HITS アルゴリズムでは、トピック “motor car company” に関するコミュニティだけが抽出されたのに対して、提案手法では、“motor car company”, “motorcycle company”, “kit car company” の3つのトピックに関するコミュニティが抽出された。トピック “motor company” に関連するコミュニティを1つだけ抽出した HITS アルゴリズムに対して、提案手法はトピック “motor company” に関連する3つのコミュニティを抽出することができたという点で優れていると言える。

5. ま と め

従来の Web コミュニティ発見手法では、トピックに関連するコミュニティを1つだけ抽出する。このため、広い意味を持つトピックに対しては、ユーザの目的とするコミュニティが抽出されなかったり、本来のトピックとは全く関係のないコミュニティが抽出されてしまうという問題点があった。

そこで、本研究では HITS アルゴリズムと Markov Cluster Algorithm (MCL) を組み合わせることにより、トピックに関連する複数の Web コミュニティを発見する手法を提案した。さまざまなトピックに対して Web コミュニティ群を発見する実験を行い、トピックに関連する複数の Web コミュニティを抽出することができることを確認した。

今後の課題としては、提案手法の出力に含まれる本

来のトピックとは関係のないコミュニティをフィルタリングする手法の提案や、実験結果をより詳細に分析するための、Web コミュニティのリンク構造を可視化するツールの開発などが挙げられる。また、提案手法における base 集合の作成方法についても検討する余地がある。現在は、root 集合に含まれる Web ページから 1 ホップ以内の距離にある Web ページ集合を base 集合としているが、これを 2 ホップに拡大した場合の実験結果について議論することも今後の課題である。

参 考 文 献

- 1) J. Kleinberg, “Authoritative Sources in a Hyperlinked Environment”, In Proceedings ACM-SIAM Symposium on Discrete Algorithms, pp.668-677 (1998)
- 2) 村田剛志, “参照の共起性に基づく Web コミュニティの発見”, 人工知能学会論文誌, Vol.16, No.3, pp.316-323 (2001)
- 3) 村田剛志, “ハイパーリンクのグラフ構造に基づく Web コミュニティの洗練”, 人工知能学会論文誌, Vol.17, No.3, pp.322-329 (2002)
- 4) K. Bharat, M. Henzinger, “Improved Algorithms for Topic Distillation in a Hyperlinked Environment”, In Proceedings ACM SIGIR Conference, pp.104-111 (1998)
- 5) S. Chakrabarti, B. Dom, P. Raghavan, S. Rajagopalan, D. Gibson, J. Kleinberg, “Automatic resource compilation by analyzing hyperlink structure and associated text”, In Proceedings of the 7th International World Wide Web Conference (1998)
- 6) 野村早恵子, 小山聡, 早水哲雄, 石田亨, “WEB コミュニティ発見のための HITS アルゴリズムの分析と改善”, 電子情報通信学会論文誌 D-I, Vol.J85-D-I, No.8, pp.741-750 (2002)
- 7) Stijn van Dongen, “Graph Clustering by Flow Simulation”, PhD thesis, University of Utrecht (2000)
- 8) <http://www.yahoo.com>
- 9) 森本寛, 黒柳奨, 岩田彰, “位相保存マッピング手法 CoPE におけるエネルギー関数の改良に関する研究”, 信学技報, NC-2004 (2005/03) (印刷中)