

機関番号：12608

研究種目：若手研究（B）

研究期間：2009～2010

課題番号：21760323

研究課題名（和文） ページランク計算に対する分散型確率アルゴリズム：
協調制御に基づくアプローチ研究課題名（英文） Distributed Randomized Algorithms for PageRank Computation:
An Approach based on Cooperative Control

研究代表者

石井 秀明 (HIDEAKI ISHII)

東京工業大学・大学院総合理工学研究科・准教授

研究者番号：50376612

研究成果の概要（和文）：

本研究では、サーチエンジンにおける検索結果を的確にランク付けするページランク (PageRank) アルゴリズムに着目し、より効率的な計算手法の確立を目指した。とくにマルチエージェント系の協調制御の観点から、分散型確率アルゴリズムを構築した。エージェント間の通信制約を考慮した場合やグラフの集約化に基づく場合等、アルゴリズムの高速化・ロバスト化を図った。他方、一般的な有効グラフ上の平均合意問題に対しても成果を得た。

研究成果の概要（英文）：

In this research project, we have focused on the PageRank algorithm for search engines, which is known to be effective in ranking search results and have developed efficient computation methods. From the viewpoint of multi-agent cooperative control, we have designed distributed algorithms based on randomization. Our achievements include more robust and fast algorithms under communication constraints among the agents and also those based on a novel technique of graph aggregation. On the other hand, we obtained useful algorithms for average consensus problems on general directed graphs.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,600,000	480,000	2,080,000
2010年度	1,100,000	330,000	1,430,000
年度			
年度			
年度			
総計	2,700,000	810,000	3,510,000

研究分野：工学

科研費の分科・細目：電気電子工学・制御工学

キーワード：制御工学，検索エンジン，分散型計算，確率アルゴリズム，合意問題

1. 研究開始当初の背景

ウェブ上の膨大な情報を検索する際に、サーチエンジンは不可欠なツールである。目的の情報に効率的に到達するには、検索結果が入力したキーワードに関連するだけでなく、的確にランク付けされる必要がある。Google における PageRank アルゴリズムは、各ウェブページの重要度を定量化およびランク付

けするためのアルゴリズムである。

PageRank アルゴリズムの特徴は、ウェブにおけるページ間のリンク構造を陽に用いる点である。その基本方針は、重要なページからリンクが張られるページはやはり重要、というものである。この問題は、ウェブ構造に関わるある確率行列の固有ベクトル計算として定式化される。PageRank アルゴリズム

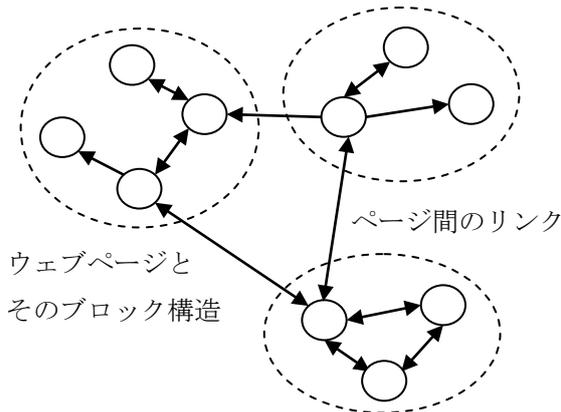


図 1：ウェブ上のマルチエージェント系

の実装時に考慮すべき重要な点は、インデックス数が 80 億ともいわれるウェブの規模である。Google では、自動的に集めた情報を基に集中的に計算を行うため、適用可能な計算手法は限定される。実際、基本的なべき乗法が用いられ、計算は 1 週間かかるとの報告がある。したがって、今後のウェブの規模拡大に対応が可能な、効率的な計算手法が求められている。

こうした背景から、研究代表者らは、最近、PageRank の計算を分散的に実行するアプローチを考え、とくに確率的な手法を活用したアルゴリズムを提案した(図 1)。ここでは、ウェブページを自律的なエージェントと見立て、ウェブ全体をマルチエージェント系として扱う。本手法は、分散的な実装が可能であり、その特徴は以下の 3 点にまとめられる。第 1 に、各ページは、自身の PageRank 値を局所的に計算可能である。このとき、リンク先および自身にリンクを張るページと各々の値を交換するための通信を行う。第 2 に、それぞれのページにおいて通信を開始するタイミングは、独立にランダムに決定される。したがって、通信のために、ページ間における順序や集中的な意思決定を必要としない。第 3 に、各ページで必要な計算量は比較的少なく、ページ間の通信量についても自由に調整可能である。

この研究について強調すべき点は、近年、制御工学やセンサネットワーク等の分野で研究が盛んなマルチエージェント合意問題により触発されたことである。興味深いことに、数学的な観点から、合意問題における基本アルゴリズムは PageRank に対する提案手法と密接な関係を持つことが明らかになっている。とくに、両者において確率行列が本質的な役割を担っており、ある種の相補的な関係が存在する。

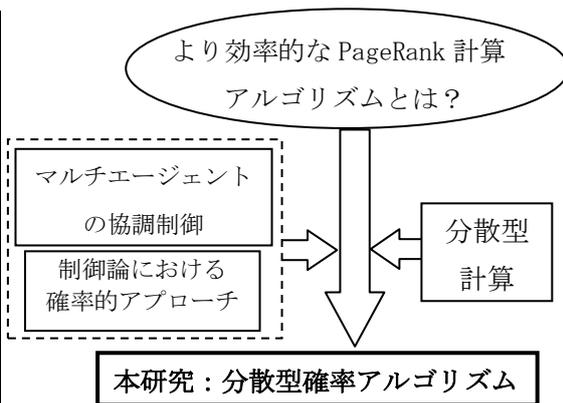


図 2：概要

2. 研究の目的

本研究では次の課題を中心に据える。まず、上記の結果で得られた基本アルゴリズムを出発点に、PageRank 計算に対する分散型確率アルゴリズムの効率化を目指し、また実用を視野に入れた機能拡張を行う(図 2)。とくに、合意問題における制御・通信アルゴリズムに関する知見やツールを最大限に利用する。それにより、従来法にはないアルゴリズム構築が可能となる。また逆に、本研究で得られる分散型計算に関する結果についても、協調制御における合意問題等への適用を考え、新たな確率アルゴリズム手法の構築や性能向上を図る。

本研究では、主に分散型確率アルゴリズムによる PageRank 値の計算手法に取り組み、とくに制御論的なアプローチをとる。具体的には、次の 5 つの課題の解決を目指す。

- (1) 基本アルゴリズムの解析：PageRank 分散型アルゴリズムと合意アルゴリズムの関係を解明する。グラフ、行列、分散計算・制御、確率アルゴリズム等の理論により多方面から解析する。
- (2) 分散型確率アルゴリズムの高速化：高精度かつ高速なアルゴリズムを構築する。制御論的な観点から、各エージェントでの計算により複雑なダイナミクスを付加する。また、エージェント間の通信における時間遅れやデータ損失等の不確かさの影響を考慮する。
- (3) 分散型確率アルゴリズムのロバスト化：現実のウェブにおけるリンクのブロック構造やリンクの更新をモデル化し、その変動に対してよりロバストなアルゴリズムを構築する。
- (4) 分散型確率アルゴリズムの総合的な性能解析：計算速度とロバスト性の間のトレードオフを明確にし、総合的な観点か

らアルゴリズムの性能限界を明らかにする。

- (5) 合意アルゴリズムに対する応用：上記の課題で得られる知見や結果を、PageRank アルゴリズムとは相補的な関係にある合意問題に対して適用する。

これらの課題に対し、主に理論的研究を進めるが、その有効性の検証は数値実験を通じて行う。また、より現実的な規模のウェブに対する PageRank 計算をシミュレーションを通じて実験する。

3. 研究の方法

本研究の目標は、研究目的で挙げた効率的な PageRank 計算のための分散型確率アルゴリズム構築を中心的な課題とする。その解決へ向けて、下記の 5 課題に段階的に取り組み、提案した基本アルゴリズムを出発点に、より高度なアルゴリズムが構成できるよう発展させていく。

- (1) 基本アルゴリズムの解析
- (2) 分散型確率アルゴリズムの高速化
- (3) 分散型確率アルゴリズムのロバスト化
- (4) 分散型確率アルゴリズムの総合的な性能解析
- (5) 合意アルゴリズムに対する応用

研究体制としては、制御論における確率アルゴリズムに詳しいトリノ工科大学のロベルト・テンポ教授を海外共同研究者として行った。本計画を通じて、短期滞在や国際会議でのディスカッションなど、これまでの共同研究の体制を強化することができた。PageRank に関わる課題(1)～(4)は主に2人で行った。また、合意問題に関する課題(5)および実験については、研究代表者とその研究室の学生が行った。

4. 研究成果

本研究で得られた成果のうち主たるものを以下にまとめる。

- (1) PageRank 計算に対する分散型確率アルゴリズム

本研究では、PageRank 計算に対して、分散型アプローチを考えるが、とくに確率的な手法を採用したアルゴリズムを提案する。本手法は、分散的な実装が可能であり、その特徴は以下の3点にまとめられる。まず、各ページは、自身の PageRank 値を局所的に計算可能である。このとき、リンク先および自身に

リンクを張るページと各々の値を交換するための通信を行う。また、それぞれのページにおいて通信を開始するタイミングは、独立にランダムに決定される。したがって、通信のためにページ間における順序や集中的な意思決定のエージェントは必要とされない。これはゴシップ型の通信方式として知られる。他方、提案した更新則においては、各ページで必要な計算量は比較的少ない。

本手法の拡張として、複数のページが同時に更新を行う場合や、各ページにおける更新停止則を実装した場合の性能解析等を行った。さらに、本アプローチに近いアルゴリズムである分散計算における非同期式反復法やマルチエージェント系の合意問題に対する更新則との相違・共通点を詳細に解析した。

- (2) 通信障害を考慮した分散型 PageRank 計算

本研究では、上記(1)で提案した分散型アルゴリズムにおいて情報論的な制約の下でのロバスト性を高めることを目指す。とくに更新を行うエージェントがリンク先の一部とのみ通信を行う状況を考える。確率的な問題設定とするために、通信先はランダムに決定されるものとする。これは例えば必要な計算や通信資源を抑制しながら更新則を実行する場合に有効であり、(1)の手法よりも自由度の高いアルゴリズムとなる。また本手法は、同時に全てのリンク先ページと情報交換ができない場合、リンク障害や通信時にデータ損失が発生する場合にも適用可能である。ネットワーク化制御や合意問題の分野における通信障害のモデルを用いて問題を定式化し、PageRank が計算可能な分散型アルゴリズムを導出した。

- (3) グラフの集約化に基づく分散型計算法

本研究では分散型 PageRank 計算に必要な計算量および通信量を抑えるために、グラフの集約化に基づく手法を提案する。分散型計算においては、各ページが自身の PageRank 値を更新するが、実際の計算はそのページをデータとして持つサーバである。そこで各サーバが、管理下のページの PageRank 値の総和を求めることを考える。このことはウェブを表すグラフを集約化することに対応し、更新すべき変数を減らすことが可能となる。提案手法では、集約化されたグラフの各ノードにおいて、上記(1)のアルゴリズムと同様にゴシップ型の通信方式に基づき分散型更新則を実行する。さらに通信に関しても、いく

つか修正を加える。まず各ノードは更新時に発リンク先とのみ通信を行う。従来法では被リンク元との情報交換も行うため、被リンク元がどのページであるか、という本来は各ページで持っていない情報を必要としていた。さらに、各ノードは発リンク先をいくつかのグループに分けて、グループ毎に通信を行う。その結果、全体の通信量を減少できるが、他方で平均の意味での収束速度についても従来法と変わらないことが示された。

(4) 一般的な有向グラフ上での量子化平均合意問題

本研究では、制御工学の分野で盛んに研究されているマルチエージェント系の合意問題を考える。とくにエージェント間の接続構造が有向グラフで表現され、通信される情報が量子化される場合を扱う。さらに通信における非同期性を実現するために、ゴシップ型と呼ばれる確率的な通信方式を採用する。量子化された情報については、各エージェントの持つ状態量を整数としてモデル化する。通常の合意問題に加えて、状態量の収束値がその初期値の平均値となることを課する平均合意問題に対する解析を行った。各々の場合について、合意を達成する分散型アルゴリズムを構築し、必要なネットワークの接続構造に関する必要十分条件を導出した。

とくに平均合意に対する条件は、既存研究におけるものよりも弱く、対称的あるいはバランスされた接続構造を要さない。これは surplus と呼ばれる付加的な変数を各エージェントに導入することで達成可能となった。提案アルゴリズムにおける収束時間に関する解析も行い、エージェント数に関して多項式時間のオーダーであることを確認した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計4件)

- ① Kai Cai and Hideaki Ishii, Quantized consensus and averaging on gossip digraphs, *IEEE Transactions on Automatic Control*, 掲載決定, 2011. 査読有
- ② Tomoyuki Kogure and Hideaki Ishii, Optimal design of cyclic pursuit weights in hierarchical multi-agent systems, *International Journal of Control*, 掲載決定, 2011. 査読有

- ③ Kai Cai and Hideaki Ishii, Quantized average consensus on gossip digraphs with reduced computation, *SICE J. Control, Measurement, and System Integration*, 掲載決定, 2011. 査読有
- ④ Hideaki Ishii and Roberto Tempo, Distributed randomized algorithms for the Page-Rank computation, *IEEE Transactions on Automatic Control*, 55: 1987-2002, 2010. 査読有

[学会発表] (計18件)

- ① Kai Cai and Hideaki Ishii, Average Consensus on General Digraphs, 第11回計測自動制御学会 制御部門大会, 2011年3月17日, 沖縄
- ② Hideaki Ishii, Roberto Tempo, and Er-Wei Bai, Distributed randomized PageRank algorithms based on web aggregation over unreliable channels, *49th IEEE Conference on Decision and Control*, 2010年12月17日, Atlanta, GA, USA
- ③ Kai Cai and Hideaki Ishii, Convergence time analysis of quantized gossip algorithms on digraphs, *49th IEEE Conference on Decision and Control*, 2010年12月17日, Atlanta, GA, USA
- ④ Tomoyuki Kogure and Hideaki Ishii, Optimal design of cyclic pursuit weights in hierarchical multi-agent systems, *49th IEEE Conference on Decision and Control*, 2010年12月17日, Atlanta, GA, USA
- ⑤ 岡野 訓尚, 石井 秀明, 原 辰次, 情報理論的アプローチによるネットワーク化制御系の感度解析, 第2回高信頼性無線制御研究会, 2010年11月1日, 名古屋
- ⑥ 石井 秀明, マルチエージェント系の協調制御と PageRank の分散型計算, JST/CRDS制御システム分科会・SICE研究会合同ワークショップ, 2010年10月5日, 東京
- ⑦ 石井 秀明, Roberto Tempo, Er-Wei Bai, 不確かな通信を介した集約ウェブに基づく分散型確率 PageRank アルゴリズム,

第 39 回制御理論シンポジウム, 2010 年 9 月 28 日, 大阪

- ⑧ Kai Cai and Hideaki Ishii, Convergence Time Analysis of Quantized Consensus Algorithms on Gossip Digraphs, 第 39 回制御理論シンポジウム, 2010 年 9 月 28 日, 大阪
- ⑨ Kai Cai and Hideaki Ishii, Further results on randomized quantized averaging: A surplus-based approach, *2010 IEEE Int. Symp. Computer-Aided Control System Design*, 2010 年 9 月 9 日, Yokohama, Japan
- ⑩ 石井 秀明, ネットワーク化制御系: 通信制約下で制御する, 第 1 回高信頼性無線制御研究会, 2010 年 7 月 28 日, 東京
- ⑪ Hideaki Ishii, Koji Tsumura, and Tomohisa Hayakawa, Coarseness in quantization for stabilization of linear systems over networks, *19th International Symposium on Mathematical Theory of Networks and Systems*, 2010 年 7 月 6 日, Budapest, Hungary
- ⑫ Kai Cai and Hideaki Ishii, Gossip consensus and averaging algorithms with quantization, *2010 American Control Conference*, 2010 年 7 月 2 日, Baltimore, MD, USA
- ⑬ Hideaki Ishii, Distributed PageRank Computation with Link Failures, *Symposium on Systems, Control, and Signal Processing*, 2010 年 3 月 31 日, 京都大学
- ⑭ Hideaki Ishii, Distributed Randomized PageRank Computation Based on Web Aggregation, *International Workshop on Networked Control Systems*, 2009 年 12 月 29 日, Shenzhen, China
- ⑮ 石井秀明, Roberto Tempo, 分散型確率アルゴリズムに基づく PageRank 計算, 計測自動制御学会 システム・情報部門学術講演会, 2009 年 11 月 24 日, 横浜市
- ⑯ 石井秀明, Roberto Tempo, 通信障害を考

慮した分散型 PageRank 計算, 計測自動制御学会 第 38 回制御理論シンポジウム, 2009 年 9 月 16 日, 大阪市

- ⑰ Kai Cai and Hideaki Ishii, Quantized Consensus on Gossip Digraphs, 計測自動制御学会 第 38 回制御理論シンポジウム, 2009 年 9 月 16 日, 大阪市
- ⑱ 木暮知行, 石井秀明, 階層型マルチェーメント系における巡回追従重みの最適設計, 計測自動制御学会 第 38 回制御理論シンポジウム, 2009 年 9 月 16 日, 大阪市

[図書] (計 1 件)

- ① Hideaki Ishii and Roberto Tempo, Distributed PageRank computation with link failures, in J. C. Willems, S. Hara, Y. Ohta, and H. Fujioka (editors), *Perspectives in Mathematical System Theory, Control, and Signal Processing*, Lecture Notes in Control and Information Sciences, vol. 398, pp. 139-150, Springer, Berlin, 2010.

[その他]

ホームページ等

<http://www.sc.dis.titech.ac.jp/ishii>

6. 研究組織

(1) 研究代表者

石井 秀明 (HIDEAKI ISHII)

東京工業大学・大学院総合理工学研究科・准教授

研究者番号: 50376612