

令和 5 年 6 月 8 日現在

機関番号：33917

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K11856

研究課題名（和文）複素数を用いたリンク解析による情報検索アルゴリズム研究

研究課題名（英文）A Study for the Search Engine Algorithm based-on Link-Analysis Using Complex Numbers

研究代表者

杉原 桂太（Sugihara, Keita）

南山大学・理工学部・講師

研究者番号：40410758

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：情報検索のための PageRank はネットワークのリンク解析によるノードへのスコア付与アルゴリズムとして広く普及している。PageRank には高計算コスト等の問題があり、改良策が盛んに研究されて来た。従来の改良は当初の PageRank と同じくネットワークの隣接行列を用いる。しかし、PageRank の問題は同行列では根本的には解決されない懸念がある。本研究では、エルミート隣接行列を出発点とし、より PageRank と同様のランキング結果をもたらす、かつ PageRank の問題点を回避した手法について検討を行なった。

研究成果の学術的意義や社会的意義

情報検索のための PageRank はネットワークのリンク解析によるノードへのスコア付与アルゴリズムとして広く普及しており、これに替わる手法の開発には社会的な意義が認められる。

研究成果の概要（英文）：PageRank is the widely used algorithm for search engines based-on link-analysis of the network. PageRank algorithm has problems such as high calculation cost. Improved PageRank algorithms have been increasingly studied. Those studies employ the adjacency matrix of the network as with PageRank. However, there is a concern that the problems of PageRank cannot be solved by the matrix. This study has utilized the Hermitian adjacency matrix to score a node of the network. By using the matrix, the study studied how a similar tendency ranking result with PageRank can be realized and how the problems of PageRank are revolved.

研究分野：エルミート隣接行列

キーワード：情報検索 PageRank エルミート行列 エルミート隣接行列

1. 研究開始当初の背景

情報検索のための PageRank はネットワークのリンク解析によるノードのスコア付与アルゴリズムとして広く普及している。PageRank には高計算コスト等の問題があり、改良策が盛んに研究されて来た。

PageRank の問題は隣接行列を用いることから発生するダンピング・ファクターという係数を活用することに由来する。隣接行列はネットワークのノードの数を行数、列数とした正方行列で、ノード u からノード v へのリンクがある場合は (u, v) 成分は 1、ノード u からノード v へのリンクがない場合には (u, v) 成分は 0 となる行列である。

従来の改良は当初の PageRank と同じくネットワークの隣接行列を用いる。しかし、PageRank の問題は同行列では根本的には解決されない懸念がある。というのは、改良策においても隣接行列を用いるゆえ、ダンピング・ファクターを使用することになるからである。

2. 研究の目的

本研究は、ダンピング・ファクターを必要とする隣接行列に替え、ダンピング・ファクターを必要としないエルミート隣接行列に依るスコア付与のアルゴリズムを開発し、PageRank を凌駕し発展性を持つ手法を構築する。

エルミート隣接行列とは、エルミート行列を用いた行列である。エルミート行列は、行列の各成分をその共役複素数で置き換えた行列の転置行列が元の行列に等しい行列である。

エルミート隣接行列はネットワークのノードの数を行数、列数としたエルミート行列で、各成分は次のように定義される。すなわち、エルミート隣接行列においては、ノード u からノード v へのリンクがあり、かつ、ノード v からノード u へのリンクがある場合は (u, v) 成分は 1、ノード u からノード v へのリンクがあり、かつ、ノード v からノード u へのリンクがない場合には (u, v) 成分は i 、ノード u からノード v へのリンクがなく、かつ、ノード v からノード u へのリンクがある場合には (u, v) 成分は $-i$ 、ノード u からノード v へのリンクがなく、かつ、ノード v からノード u へのリンクがない場合には (u, v) 成分は 0 となる。ここでエルミート隣接行列の定義において i は虚数単位を表す。

3. 研究の方法

エルミート隣接行列を用いてスコアをノードに付与するアルゴリズムを考案し、アルゴリズムに基づく手法をプログラミング化する。計算機において、エルミート隣接行列によるアルゴリズムと PageRank からなるアルゴリズムを同一のネットワークに適用し、各アルゴリズムによるノードのスコア・ランキングと各アルゴリズムの計算時間を比較する。

4. 研究成果

エルミート隣接行列を用いるアルゴリズムについて、PageRank と比較して次のことが明らかになった。PageRank においては、高スコアのノードからのイン・リンクを多数持つノードのスコアが高くなる。PageRank ではノードからのアウト・リンクは当該のノードのスコアには影響しない。

これに対してエルミート隣接行列に依るアルゴリズムでは、エルミート行列の性質から、ノードからのイン・リンクとアウト・リンクの双方が当該のノードのスコアを高めることになる。これは、エルミート隣接行列を用いる手法においては、複素平面上において、エルミート隣接行列によるノードに対応する点について実軸からの角度を用いるためである。

エルミート隣接行列によるアルゴリズムについては明らかになった以上のことから、PageRank と同様の傾向のノードのスコア・ランキングを実現するためには以下の手法が必要であることが分かった。

この手法では、ネットワークに次の 2 つのノードを加える。1 番目のノードは、ネットワークの全てのノードへのアウト・リンクを持つノードである。2 番目のノードは、ネットワークの全てのノードからのイン・リンクを持つノードとなる。これら 2 つのノードを加えたネットワークにおいて各ノードのスコアについてエルミート隣接行列を用いて計算すると、PageRank によるスコアと同じ傾向のノード・スコアを得ることができる。

本研究では、こうした手法を用いて、PageRank と比較してノードのスコアに必要な計算時間を抑えたアルゴリズムを開発した。

具体的には、ノードの数が 60 のネットワークを構成し、PageRank によるノード・スコアのランキングと本研究によるノード・ランキングとの間でスピアマンの順位相関係数を求めたところ、0.9752789 という高い値が得られた。

さらに、ノードの数が 2000 までの 20 のネットワークをランダムに発生させ、各ネットワークのノードのスコアを算出するために必要な時間を PageRank と本研究による手法の間で比較した。その結果、何のネットワークにおいても本研究によるアルゴリズムの方が計算時間が短く、その差はノードの数が増えるほど大きくなることが分かった。

ここから得られた成果から、さらなる研究の必要性として、よりノードの数の多い大規模なネットワークにおいて同様の結果が得られることを示すことが挙げられる。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 杉原 桂太	4. 巻 21
2. 論文標題 A Novel Node Ranking Algorithm Using Complex Numbers : A Comparison to Google PageRank	5. 発行年 2021年
3. 雑誌名 アカデミア. 理工学編 : 南山大学紀要	6. 最初と最後の頁 46 - 53
掲載論文のDOI (デジタルオブジェクト識別子) 10.15119/00003082	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計5件（うち招待講演 0件/うち国際学会 4件）

1. 発表者名 Zhenguo Hu, Hirokazu Hasegawa, Yukiko Yamaguchi, and Hajime Shimada
2. 発表標題 High-Performance Distributed NIDS Cluster Based on Hybrid Detection Platform
3. 学会等名 情報科学技術フォーラム FIT 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 Zhenguo Hu, Hirokazu Hasegawa, Yukiko Yamaguchi, and Hajime Shimada
2. 発表標題 Towards Network-Wide Malicious Traffic Detection with Power-Effective Hardware NIDS Design (Poster)
3. 学会等名 the 25th IEEE Symposium on Low-Power and High-Speed Chips (COOLChips 25) (国際学会)
4. 発表年 2022年

1. 発表者名 Keita Sugihara
2. 発表標題 Calculation Cost Efficient Alternative to Google PageRank
3. 学会等名 電子情報通信学会2021年総合大会
4. 発表年 2021年

1. 発表者名 Keita Sugihara
2. 発表標題 Tractable and Calculation cost efficient Alternative to Google PageRank
3. 学会等名 WAW 2020 17th Workshop on Algorithms and Models for the Web Graph (国際学会)
4. 発表年 2020年

1. 発表者名 Keita Sugihara
2. 発表標題 Beyond Google 's PageRank: A Novel Link Analysis Algorithm without a Damping Factor
3. 学会等名 International Network for Social Network Analysis (国際学会)
4. 発表年 2020年

〔図書〕 計0件

〔出願〕 計0件

〔取得〕 計1件

産業財産権の名称 情報処理システム及び情報処理方法	発明者 杉原桂太	権利者 同左
産業財産権の種類、番号 特許、第6919961号	取得年 2021年	国内・外国の別 国内

〔その他〕

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	嶋田 創 (Shimada Hajime) (60377851)	名古屋大学・情報基盤センター・准教授 (13901)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------