

著者論文行列ランク APM を用いた論文検索結果からの 著者ランキング

大阪欣吾[†] 黄 宏軒[†] 川越恭二[†]

[†]立命館大学 情報理工学部

1 はじめに

研究者が研究を行うには、研究分野の重要な研究者の論文の調査が必要となる。現在、論文情報や著者情報が網羅的に収集され、学術論文情報検索のサービスの提供が行われている。日本国内では、国立情報研究所が提供する NII 論文情報ナビゲータ Ci N ii¹がある。海外では、Web of Science²などのサービスがある。

しかし、既存サービスには2点問題が存在する。まず、利用者が入力した文字列が含まれていないと論文・著者の検索できないことである。次に、検索キーワード・分野に関する重要な著者名を調べるシステムは存在しないことである。

そこで、論文情報と論文著者情報に着目し、著者論文行列 Autor and Paper Matrix(以下、APM) ランクを用いた論文検索結果からの著者ランキングを本研究で提案する。ここで、論文情報とは、論文の参考文献、引用文献の情報とする。論文著者情報とは、論文における著者情報とする。論文情報を用いることでキーワードでは取得できなかった論文・著者を取得でき、さらに、論文著者情報を基に重み付けを行い、影響力の高い著者を取得できる。これらの情報を基に行列を生成し、その行列を著者論文行列 APM と定義する。APM ランクを用いることで、検索キーワード・分野に対しての重要な研究者の論文を検索することが可能となる。

2 著者論文行列ランクを用いた著者ランキング手法

2.1 基本的考え方

著者論文行列 APM は、グラフ理論を基に生成される行列とする。ノードは、著者と論文の情報を1ノ

ードとし、エッジはノードからノードへのリンク関係とした。エッジの重みがあり、論文における影響力が対応しており、影響力を降順に「筆頭著者、末尾著者、その他の著者」と定義する。そのうえ、以下の4点を考慮してエッジの重みを決定する。

1. Lead Author を論文において実際的な貢献を持っている者と定義する。このことから、筆頭著者は当該論文に最も影響力をもつであろうことから [2]、本研究では筆頭著者を Lead Author にする。
2. Senior Author を当該論文の監督・責任者と定義する。その分野に関する知識所持者と考え、Lead Author の次に Senior Author が影響力を持っている者とする。末尾著者はこの役割を担うことが多いことから [1]、本研究では末尾著者を Senior Author にする。
3. 1つの論文内で1人の著者が2種類以上の役割を兼任することはないとする。
4. 自著論文を参考文献として用いる場合は、他人の論文を参考文献にしている場合では、影響力は低いとする。なぜなら、他の研究者に有効な情報を与える論文という観点からは、実質的に意味が無いからである。

上記の重み付けエッジを持つグラフを基に算出した隣接行列を APM とする。APM に対して PageRank [3] の計算を行う。PageRank には、高い評価を持つノードからリンクを受けているノードは評価が高いという特徴があり、このことは、論文の影響力にも使えると考えた。計算の結果、すなわち、各論文から算出されたキーワード・分野へある著者名が与える影響度を著者名ごとに合計し、ソートすることで論文著者情報をランク付けが可能となる。その結果、検索キーワード・分野に対する影響力・専門性の高い著者検索が可能となる。

Author Ranking from Paper Search Results using Author and Paper Matrix Rank
Kingo OSAKA[†], Hung-Hsuan HUANG[†] and Kyoji KAWAGOE[†]

[†]College of Information Science and Engineering, Ritsumeikan University.

{osaka@coms.ics, huang@fc, kawagoe@is}.ritsumei.ac.jp

¹<http://ci.nii.ac.jp/>

²<http://apps.webofknowledge.com/>

2.2 提案手法の処理手順

本節では、著者論文行列 APM ランクを用いた論文検索結果からの著者ランキング手法の処理を述べる。

STEP1 ノード生成

本研究では、著者と論文の情報を1ノードと定義する。1ノードと定義することによって、著者と論文の2つの情報を考したランキング付けを可能とする。

STEP2 エッジ生成

STEP1 で生成したノードを基に以下の事を行う。

- 同一論文の著者の場ありにその著者間にエッジを生成する
- ある論文とその論文の参考文献に関してその論文の筆頭著者から参考文献の著者にエッジを生成する
- ある論文とその論文の引用文献に関してその論文の筆頭著者から被引用文献の著者にエッジを生成する

生成されたノード・エッジを基にしたグラフを隣接行列に変換する。まず、隣接行列 E を生成する。ここで行はある論文 P_i : 著者 A_j ノードとし、列はある論文 P_k : 著者 A_l ノードとする。 E の成分 $e_{(A_j P_i \rightarrow A_l P_k)}$ は、行から列にエッジがある場合には1とし、そうでない場合を0とする。

STEP3 エッジに対して重み付け

本研究では以下の2点の重み付けを行う。ここで、主論文とはキーワード検索で取得できた論文と定義する。

共著論文著者順の重み付け ω_c

- 筆頭著者に流入してくるエッジの場合 : 3
- 末尾筆者に流入してくるエッジの場合 : 2
- その他著者に流入してくるエッジの場合 : 1

自著論文の重み付け ω_s

- 主論文の筆頭著者と参考文献の筆頭著者が同じノードへのエッジの場合 : 0
- 主論文の著者名と参考文献の著者名が同じノードへのエッジの場合 : 1/2
- 主論文の著者名と参考文献の著者名が異なるノードへのエッジの場合 : 1

上記の重みを E に当てはめ、重み付き隣接行列 R を生成する。ここで R の成分 $r_{(A_j P_i \rightarrow A_l P_k)}$ を (1) 式で示す。

$$r_{(A_j P_i \rightarrow A_l P_k)} = e_{(A_j P_i \rightarrow A_l P_k)} * \omega_c * \omega_s \quad (1)$$

論文著者ノード影響度

P1:A1	0.005715
P1:A2	0.002735
P1:A3	0.002735
P1:A4	0.004254
P2:A5	0.002418
...	...

著者別影響度

1	A4	0.096675
2	A2	0.035594
3	A96	0.031001
4	A31	0.023136
5	A70	0.019208
...

図 1: APM 行列計算結果及び著者ランク結果

STEP4 ランキングの計算

STEP3 で生成された重み付き隣接行列 R を用いて、PageRank の計算を行う。その計算結果、すなわち、各論文から算出されたキーワード・分野へある著者名が与える影響度を著者名ごとに合計し、ソートすることで論文著者情報をランク付けする。

3 APM を用いた計算例

「WEBDB2008Web 検索とランキング・2010 検索支援・2011 検索とランキング」の論文情報を基に著者論文行列 APM を用いたランク付けを行った。計算結果の例を図 1 に示す。なお、便宜上論文名及び著者名をアルファベットに置き換えた。図 1 の左の表は、論文著者ノードの影響度を表している。図 1 の右の表は、左の表を基に求めた著者別影響度を表している。この結果、著者 A4 が最も影響度の高い重要な研究者であることを示している。

4 終わりに

今後、著者論文行列 APM ランクを用いた論文検索結果からの著者ランキングの評価を行う。更に、著者間のリンク関係だけではなく収録刊行物で論文の重み付けを行うことによる改善を目指す。

参考文献

- [1] William M.: "Responsibilities of Authorship", CHEST December 2007 vol.132
- [2] International Committee of Medical Journal Editors. (http://www.icmje.org/urm_full.pdf), 2011
- [3] L. Page, et al.: The PageRank Citation Ranking: Bringing Order to the Web, Stanford Digital Libraries Working Paper, 1998