

ページランク・アルゴリズムの可視化 —テレポーテーションの秘密—

安田 友希 田嶋 祐衣 向 直人
 相山女学園大学 文化情報学部 文化情報学科

1. はじめに

検索エンジンの中でも多くのシェアを誇るのが Google である。その理由は、Google の検索エンジンの中核である、ページランク・アルゴリズムが利用者の求めるウェブページを高い精度でランク付けして提供するからである。しかし、我々の多くはその仕組みを知らないまま利用している。そこで本研究では、ページランク・アルゴリズムの仕組みを aritsoc で可視化し、理解を深めることを目的とする。

2. ページランクの基本公式

ウェブページ間のリンクをウェブページの推薦と捉え、多くのウェブページからリンクされたウェブページの評価を高く設定する。ここでは、図 1 に示すネットワークを考える。矢印がリンクを表し、ウェブページ P_j からウェブページ P_i に遷移するリンクが存在することを意味している。



図 1 ウェブページ間のリンク

この場合、このリンクを、ウェブページ P_i の“入リンク”、ウェブページ P_j の“出リンク”と呼ぶ。ここで、ウェブページ P_i のページランク $r(P_i)$ は式(1)で計算される。 $|P_j|$ はウェブページ P_j からの出リンクの総数を表し、 B_{P_i} はウェブページ P_i に対し入リンクを持つウェブページの集合である。ウェブページ P_i のページランク $r(P_i)$ がリンクを経由して他のウェブページに拡散していくことが分かる。

$$r(P_i) = \sum_{P_j \in B_{P_i}} \frac{r(P_j)}{|P_j|} \quad \text{式(1)}$$

3. Artisoc によるページランクの視覚化

エージェントによるページランクの伝播の様子を図 2 に示す。今、 P_1 に 10 体のエージェントが存在する。これらのエージェントは P_1 のリンクを辿り、他のウェブページに移動する。このとき、エージェントは P_1 のページランク $r(P_1)$ を、10 体で分け合い、移動先のウェブページに運ぶ。例えば、 $r(P_1) = 1$ とすると、1 体のエージェントが持つランクは $1/10$ となる。 P_1 の出リンクは P_2 と P_3 の 2 つであるため、エージェントは 50% の確率で P_2 または P_3 に移動する。この結果、 P_2 と P_3 にそれぞれ 5 体のエージェントが移動したとすると、 P_2 と P_3 のページランクは $5/10 = 1/2$ となる。

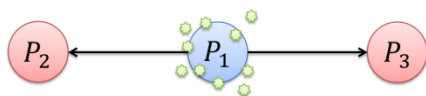


図 2 エージェントによるページランクの伝播

4. エージェント

エージェントのフローチャートを図 3 に示す。まず、エージェントは、現在訪れているウェブページから、他のウェブページへ運搬するページランクの量を計算する。次に、ウェブページが出リンクを有しているかどうかを調べる。出リンクを有している場合は、リンクを辿り、移動先のウェブページに運び出したページランクを加算する。しかし、「閉じ込め」や「閉路」がある場合、ページランクが閉じたネットワーク内に溜まってしまう。そこで、登場するのが“テレポーテーション”という概念である。ページランク・アルゴリズムにおけるテレポーテーションとは、一定の確率でリンクの有無とは無関係に、他のウェブページにページランクを伝播することを意味する。6 つのウェブページで構成されるネットワークにエージェントを適用した結果が図 4 である。全てのウェブページのページランクの値が“0”になることなく、一定の値に収束していることが分かる。

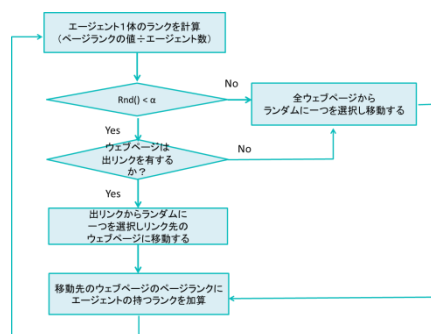


図 3 フローチャート

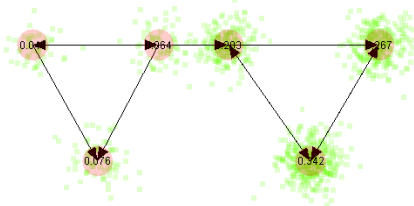


図 4 ページランクの結果

5. まとめ

本研究では、誰もが利用する検索エンジンの裏側に潜むページランクというアルゴリズムに焦点を当て、本来は行列計算が必要な導出手続きを、エージェントモデルに基づき実装することで、可視化することに成功した。これにより、数学者などの専門家だけでなく、情報工学を学ぶ学生はもちろん、インターネットを利用する全ての人々に、ページランクの仕組みを直感的に伝えることが可能になった。