

情報検索の手法を応用した設問の重要度によるランキング手法

飯田 宗玄
東京都立産業技術高専専攻科創造工学専攻

保福 一郎
東京都立産業技術高専ものづくり工学科

1 はじめに

設問に対し難易度による評価を行う際は様々な手法が考えられる。例えば経験に基づいたもの(試験の配点など)や、受験者の正答率などの情報を利用するものなどがあるが、しかし、前者には評価する者の主観が多分に影響すること、後者には事前(試験を行う前)に利用できないなどの欠点がある。

そこで我々は、事前に得られる情報から設問を評価する尺度を提案する。この手法では、各設問で適用されている定理・定義の情報と、各定理・定義間の関係性のみが利用され、これらの情報から設問間の類似性を表す行列を生成し、この行列に PageRank アルゴリズムを適用し設問の重要度を算出する。PageRank とは Google による Web 上の情報検索で利用されているランキング手法である ([1], [2])。

さらに、実際の試験問題に適用したときの結果も併せて示す。

2 設問の評価

PageRank ではページ間に張られたリンクの構造から「重要度」と呼ばれる数値を算出し、ランク付けを行う。重要度は、そのページに向かって張られたリンクを持つすべてのページの重要度をもとに再帰的に計算される。

設問に適用する場合は設問間の関係を定義しなければならないが、それは主観に頼らなければならない、設問数が増えていけば非常に複雑な作業になってしまう。そこで、本手法では、設問間の関係性は直接与えず、定義および定理の関係と、各設問に対しどの定義または定理が関連しているかの関係を与えてランキングを行う。また、本手法では設問、定義および定理が同一の尺度により評価される。

設問、定義または定理 A_i に対する重要度を $r(A_i)$ で表し、すべての重要度をならべたベクトルを

$$\pi = (r(A_1), r(A_2), r(A_3), \dots, r(A_N))^T \quad (1)$$

と書くことにする。最終的にランキングを行うにはこのベクトルの要素同士を比較して昇順に並べ直せ

ばよい。

重要度を求めるために、まず設問、定義および定理の相互関係を有向グラフと捉え、隣接行列 $P_{\text{raw}} = \{pr_{ij}\}$ を以下のように定義する。

$$pr_{ij} = \begin{cases} 1 & T_j \text{ に } T_i \text{ が適用されている} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

この行列の、各行は定理、定義または設問を表し、各列は定理または定義を表す。ただし定理、定義、設問を特定せずに T_i と書いた。

つぎに、これと転置行列との積 $P_{\text{raw}} P_{\text{raw}}^T$ をとり、各列を正規化した行列 $P' = P_{\text{raw}} P_{\text{raw}}^T$ を使い、以下のようにして、「マルコフ過程」の推移確率行列の形式を取る行列をえる。

$$P = (1 - c) \frac{1}{N} E + c P' \quad (3)$$

ただしこの「過程」はもとの行列 P_{raw} の表すグラフとは別のグラフで書かれる。なお、減衰係数 c ($0 < c < 1$, ふつう 0.85 程度に設定される) を導入して行列の零要素をなくし、この後の計算を確実に収束させるといふ工夫をしているが、これは PageRank 由来のものである。

以後の計算は PageRank とまったく同じ手順となり、この P を使って

$$\pi_j^T = \pi_{(j-1)}^T P \quad (4)$$

のように漸化式で表される計算を j 回くり返して、 $j \rightarrow \text{inf}$ の極限では $\pi_j \rightarrow \pi$ とする π を(数値計算によって近似的に)求める。なお、 P は各列の合計が 1 になるようにしてあるから、絶対値最大の固有値は 1 であり、つまり、この操作はべき乗法により P の絶対値最大の固有値に対する固有ベクトル π を求めることに等しい(ペロン-フロベニウスの定理)。

3 実際の試験問題へ適用した例

ここまでで説明した設問の評価方法を、実際の試験問題(高専 1 年生むけの授業、「代数」の定期考查問題)に適用した結果を示し、どのような性質のものなのか説明を試みたい。

試験の出題範囲は「三角関数の基礎」である。この範囲に含まれる定理および定義とそれらの相互関係は、図1の有向グラフのとおりになっていると仮定する。また、各設問と、それを解く際に必要になる定理または定義の関連は、表1のようにになっているものとした。

これら2つをまとめて隣接行列のように表現すると

$$P_{\text{raw}} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad (5)$$

これから推移確率行列 P を作成し、重要度 π を計算して設問、定理または定義をランク付けすると表2のようになる。

4 結論

設問を評価する尺度(重要度)を得るために、PageRank アルゴリズムを利用する方法を提案した。また、どのようなランキングが作成されるのかを、例により示した。

参考文献

[1] S. Brin, L. Page “The Anatomy of A Large-scale Hypertextual Web Search Engine”, Proceedings

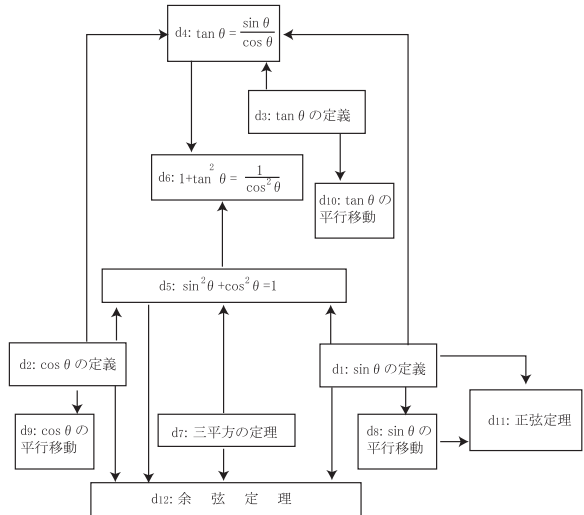


図1 定理・定義の相互関係

表1 設問と定理または定義の関連

Q	d											
	1	2	3	4	5	6	7	8	9	10	11	12
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												
12												

表2 ランク付けした結果

ランク	1	2	3	4	5	6	7	8	9
問題	12	9	8	6	3	10	18	17	4
10	11	12	13	14	15	16	17	18	19
1	2	7	11	13	5	20	14	19	16

of the seventh international conference on World Wide Web 7, p.107–117 (1998).

[2] A. N. Langville and C. D. Meyer “A Survey of Eigenvector Methods for Information Retrieval”, SIAM Review Vol.47 No.1, pp.135–161 (2005).