

2. 論文誌 i を引用する論文誌の集合を $In(i)$ とする.
3. 論文誌 i が引用する論文誌の集合を $Out(i)$ とする.
4. $In(i) \cap Out(i)$ の要素数がしきい値以上の場合, i をリンクファームの疑いがある論文誌として登録する.
5. 全ての $i \in \{1, \dots, N\}$ について以上を行う.
6. 疑わしいとされた論文誌の集合を SeedSet とする.

● SeedSet の拡張

1. 論文誌 i が引用する論文誌の集合を $Out(i)$ とする.
2. $Out(i) \cap \text{SeedSet}$ の要素数がしきい値以上なら論文誌 i を SeedSet に追加する.
3. 全ての $i \in \{1, \dots, N\}$ について以上を繰り返す. SeedSet が変化しなくなったら終了.

● リンクファームへのペナルティ

論文誌 i が SeedSet に属する論文誌 j を引用している場合,

– $A_{ij} = 0$ とする.

または

– 論文誌 i が SeedSet に属する論文誌を n 冊引用している場合, $A_{ij} = 1/n$ とする.

「SeedSet の検出」はリンクファームを形成する論文誌群は相互に引用することを利用している. 「SeedSet の拡張」は SeedSet に属する論文誌を引用する論文誌も疑わしい可能性があることを利用している. 最終的に得られた SeedSet がリンクファームに属する論文誌の集合を表す. 「リンクファームへのペナルティ」については, リンクファームに属する論文誌の EF を直接下げる, などこの他の方法も考えられる.

実際にこの手法を試してみることは行っていない. ただ, EF と PageRank アルゴリズムの類似性により, SEO スпам対策のアルゴリズムが論文誌の評価において現れる種々の問題に対して有効であると期待できそうである. EF の他の問題点についてもこのようなアナロジーが有効かどうかは引き続き調査していきたい.

4

4.1

参考文献

- ¹M. Franceschet, “Ten good reasons to use the eigenfactor and tm metrics”, *Inf. Process. Manage.* **46**, 555–558 (2010).
- ²増田 直紀, “アイゲンファクターを知る”, *統計数理* **61**, 147–166 (2013).
- ³B. Wu and B. D. Davison, “Identifying link farm spam pages”, in Special interest tracks and posters of the 14th international conference on world wide web (2005), pp. 820–829.