

セレンディピティ向きのクラスタリング

山崎 亜梨紗[†]塩谷 勇^{††}

† 法政大学理工学部創生科学科

†† 法政大学理工学研究科

1. はじめに

本報告ではセレンディピティ(SerenDipity, SD)向きのクラスタリングを提案する。従来の SD 指向の検索は、ネット上の情報を効率よく探し、主に情報推薦に基づいている。例えば、ネットショッピングで、A 商品を購入した人は B 商品も購入していると知らせてくれる。確かに、これが必要なかと背景にあるものを推測して必要性を認識する。これも、新しい知識で、新しい発見である。「おむつを購入する人はビールも購入するのか」と背後にあるものを推測して認識する。場合により、理由をネットで検索する。SD を利用した検索は、情報推薦に基づくものばかりでない。SD に基づいた情報検索をどのように考えるかによって、結果が異なる。セレンディピティは偶然とサガシティ(察知力, sagacity)という研究も報告されている[8]。インターネット上の情報は、人類が作り出した最も巨大な知識であるならば、単に情報を探すだけでなく、検索の利用者に新しい発見となる情報を積極的に支援したい。究極的には、情報を入力したら、新しい発見を計算機がする。

本報告では、[5]を研究の動機に、複雑ネットワークの特性を利用したクラスタリング手法を提案する。また、選択の科学[1]と弱い紐帯[2]に基づいたクラスタリング手法を提案する[4]。選択の科学は簡単に言えば、選択肢が多いほど利用者の選択肢が広がり良いと考えられるが、利用者にとってどれを選択して良いか判断に迷い、結局は選択できない。検索で言えば、検索ランキングのトップ3を除けば、多くの利用者は訪れることがない。サーチエンジンの最適化が必要な理由はこの点にある。これは一般論であるが、専門家になると、選択肢が多い方が良いことが知られている。ネットワークの情報は、大量で、専門家といえども、情報により限度を超えている。

弱い紐帯は、友人を頼りに仕事を探すためには、誰に聞くのが良いかという問題で知られている。頻繁に連絡しあう友人か、または、しばらく連絡をしていない友人かである。「しばらく連絡をしていない友人」という結果が得られている。これは、頻繁に連絡しあう仲間ならば、多くの情報を共有しており、Axelrod の文化の感染[3]のように特性が似ており、価値のある新しい情報を持っている可能性が低いことを意味する。本報告では、以上の二つの研究結果を踏まえて、SD 向きのクラスタリング手法の提案によって、新しい検索手法を提案する。毎日新聞の記事の特性に基づく有効性を確認する。2020 年の 1 月の毎日新聞の一部の記事

の語は出現頻度を ChaSen[7]により形態素解析して出現語を手作業で構成し結果[6]をジフの法則に基づいて語を選択した。各記事の特性は出現語の特性とし、次の述べる SD 向きのクラスタリングを行った。

2. SD 向きのクラスタリング

従来のクラスタリングは類似度の高いオブジェクトをグループにして分類する一方、手法[4]はグループの満足度の高いものをボトムアップにグループ化する。従来手法に比較して偶然を考慮して初期値依存性が高く、満足度によるグループの質を考慮してグループ化する。

各オブジェクトの特性が一樣にランダムな時の結果が[4]に示されている。本報告では毎日新聞の記事の語の出現頻度が人間活動で構築された特性とし、各記事 ID をエージェントと考えてクラスタリングを行う。

構築されたクラスタ間の類似度からネットワークグラフを構築し、弱い紐帯の強さ[2]に基づいて重みづけをして、選択の科学[1]に基づいて適度の数のクラスタを選択する。新聞記事の語の出現頻度を新聞記事の特性としてするが、出現頻度に揺らぎを加えることで、より満足度が一時的に高くなることも利用する。

まとめとして、本報告では SD 向けのクラスタリング手法を提案し、その考察及び、手の詳細と結果については、ポスター発表にて報告を行う。

参考文献

- [1] シーナ・アイエンガー, 選択の科学. 文春文庫, 2014.
- [2] M.S. Granovetter, The Strength of Weak Ties, American Journal of Sociology, 78, 6, 1360-1380, 1973.
- [3] Robert Marshall Axelrod, The Dissemination of Culture: A Model with Local Convergence and Global Polarization, Journal of Conflict Resolution, 41, 2, 203-226, 1997.
- [4] Kodai Shinkawa and Isamu Shioya, Chaos Behavior in Group Formation, International Journal of Computer Theory and Engineering (IJCTE), 11, 2, 27-30, 2019.
- [5] 平林,塩谷,セレンディピティによるWeb サイトからの知識発見, IEICE, student poster, ISS-A-017, 2021.
- [6] 濱村,三浦, 毎日新聞コーパス構築, 2021.
- [7] ChaSen, 2.3.3, <https://chasen-legacy.osdn.jp/>
- [8] 澤泉,片井,セレンディピティの探求—その活用と重層性思考, 角川学芸ブックス, 2007.
- [9] 北山,塩谷,集団形成の実験的考察, IEICE, student-poster, 2022 発表予定.