

# セレンディピティによる Web サイトからの知識発見

平林 瑞紀<sup>†</sup>

Mizuki Hirabayashi

塩谷 勇<sup>†</sup>

Isamu Shioya

<sup>†</sup> 法政大学工学部

Faculty of Science and Engineering, Hosei University

## 1. セレンディピティとは

思いがけない出会いや偶然の発見から探していたものとは別の重要なものを見つけること、またそれに気づくことをセレンディピティと呼ぶ。科学史上で重要な発見は多くのセレンディピティが関与しているとされており、研究者によく口にされる言葉である。

## 2. エラー生成型のセレンディピティ

セレンディピティにはいくつかのパターンが存在し、他と異なる特異な点やエラーに起因する事象はエラー生成型に分類される。本研究ではエラー生成型の例から、セレンディピティは以下の様に表されると仮説を立て、検証を行う。

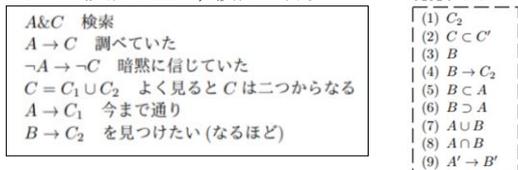


図1.セレンディピティの仮説

## 3. セレンディピティによる Web サイトからの知識発見

進路の選択など目的のあいまいな漠然とした検索は、事前知識がない状態で膨大なデータから手探りで探す必要があり、非常に困難である。そこで、検索対象の分布の違いに注目する。

例えば大学の受験先を探す場合、心理学を学ぶことができる学科にも正規分布の様に学ぶ学科もあれば、広い範囲を浅く学ぶ学科、ある特定の分野に特化した学科など様々である。そしてまた、検索者の目的や意図も様々である。検索の目的に近い分布の学科を見つけることであいまいな検索をより効率よく実現し、またそれがセレンディピティにつながるのではないかと考える。

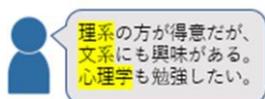


図2.進路の検索の例

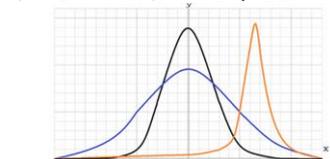


図3.心理学を学ぶ学科の予測される分布

## 4. 研究概要とアプローチ

- (1) 各学科公式紹介サイトをスクレイピングし、使用単語とその頻度を集計する。
- (2) 検索者の発言から検索単語を抽出し、共起語を補いグループ化する。グループの単語集合を検索の仮説とする。
- (3) 検索の仮説と学科サイト使用単語との紐帯を辿り、クラス候補の提案を行う。

## ” The strength of weak ties ”

弱い紐帯の強さとは、強い紐帯より弱い紐帯からの方が有益な情報が得られる可能性が高いという社会ネットワーク概念である。学科サイトの使用単語が zipf’s law に従うことを利用し、弱い紐帯と強い紐帯に分類する。仮説の理論に従って単語とクラスの繋がりを辿ることでセレンディピティを含む候補の提案が期待される。

発言者A	抽出語	共起語	頻度
強い紐帯	英語	言語	243
	心理、心、人	心理学	231
	勉強、高校	学習、教育	127
弱い紐帯	数学	数理、科学	61
	理系、文系		46
	物理	物理学、力学	42
	先生	教師	12

	全体	各クラス
部分	強い紐帯	強い紐帯
拡張	強い紐帯	弱い紐帯
否定	弱い紐帯	弱い紐帯
A'	弱い紐帯	強い紐帯

図4.単語集合グループ化の例

図5.単語とクラスの紐帯を辿る候補の提案

## 5. 結果・考察

NaiveBayes 手法をもとに、比較した結果を抜粋して紹介する。

部分	心理、心	( $\times 10^{-7}$ )
拡張	現代福祉学部	21.96846
	心理学部	2.768148
	英文学科	1.493028
	哲学科	1.461626
	史学科	1.220531
	地理学科	1.12948
	日本文学科	0.976156
	経営学部	0.266206
	応用植物科学科	0.175334
	国際文化学部	0.161588

図6. “心理、心”検索結果

A'	文系、理系	( $\times 10^{-7}$ )
否定	衛生科学科	14.27243
	経営システム工学科	0.577101
	建築学科	0.174482
	応用情報工学科	0.145999
	情報科学部	0.107698
	人間環境学部	0.095206
	地理学科	0.056474

図7. “文系、理系”検索結果

強い紐帯に分類された仮説”心理、心”に対する[部分]の候補上位2つは心理学を専門に扱う適合性の高い学科が挙がったが、英文学科があげられるという意外性も見られた。[拡張]では、文理にとらわれず様々なクラスが候補として挙げられ、意外性・発見性に富んだ提案であったと感じられる。弱い紐帯に分類された仮説”文系、理系”に対する[A']の候補はいずれも文理に関する特殊な特徴を持つクラスを拾うことができた。[否定]では、理系クラス候補が多い中で文系に分類される理系と弱い繋がりのある候補を拾うことができた。以上から、セレンディピティを考慮した学科の検索は効果的な提案が期待されると考えられる。

## 参考文献

- [1] Xuanhui Wang, ChengXiang Zhai, University of Illinois at Urbana-Champaign, Mining Term Association Patterns from Search Logs for Effective Query Reformulation, CIKM, 479-488, 2008.
- [2] 清水拓也, 土方嘉徳, 西田正吾, 発見性を考慮した協調フィルタリングアルゴリズムに関する複数方式の検討, PSJ, 2006-DBS-139, 2006.
- [3] 奥健太, セレンディピティ指向情報推薦の研究動向、知識と情報、25, 1, 2-10, 2013.
- [4] 外山滋比古 思考の整理学 1986
- [5] ABC 殺人事件 アガサ・クリスティー