

Linked Data を用いた情報推薦への取り組み

一瀬詩織[†] 大西可奈子[‡] 小林一郎[‡]

[†]お茶の水女子大学理学部情報科学科

[‡]お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース

1 はじめに

本研究では、Semantic Web 技術の枠組みの一つとして、近年盛んに開発されてきている Linked Data を使い、通常の情報推薦とは異なる、知識に基づく情報推薦を行うことを目的とする。提案する情報推薦では、Linked Data から適切な知識を抽出する手法として、DBpedia を用い、着目対象に対してリンク解析を行うことにより、対象と他のリソースとの関連性を分析する。これにより、知識間の様々な関連性を定量的に評価することができ、以前とは異なる意外な情報や関連知識に基づいた情報推薦を行えるようにする。また提案する手法の具体的な応用として、作家間の影響関係を用いた、作家の推薦を実現する。

2 Linked Data を利用した情報推薦

推薦は2つのステップに分けて行う。まず着目対象に関連するリソースを推薦候補として Linked Data より抽出し、次に、これらのリソースの重要度を、リンク解析によって求める。その結果において、スコアの高いものを着目対象に対しての推薦とする。

DBpedia 内の着目作家のリソース R から、関連する作家のリソース x を取得するため、DBpedia のエンドポイント^{*1}に対し、以下の SPARQL^{*2}クエリを用いた。

```
SELECT * WHERE {
  { <作家リソース R> ?property ?x .}
  UNION
  { ?x ?property <作家リソース R> .}
  ?x <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/ontology/Writer> .}
```

2.1 被験者実験による有効な推薦情報の発見

リソースが持つ関係は、作家によっては数十種類存在する。全ての関係を用いた場合、関連するリソー

An Approach to Information Recommendation with Linked Data

[†]Shiori ICHINOSE(g0820504@is.ocha.ac.jp),

[‡]Kanako ONISHI(onishi.kanako@is.ocha.ac.jp),

[‡]Ichiro KOBAYASHI(koba@is.ocha.ac.jp)

[†]Dept. of Information Sciences, Faculty of Science, Ochanomizu University,2-1-1 Ohtsuka Bunkyo-ku Tokyo 112-8610

[‡]Advanced Sciences, Graduate School of Humanities and Sciences, Ochanomizu University,2-1-1 Ohtsuka Bunkyo-ku Tokyo 112-8610

^{*1}<http://dbpedia.org/sparql>

^{*2}<http://www.w3.org/TR/rdf-sparql-query/>

スの数は膨大になるため、今回は、推薦対象の抽出に利用する関係を、特定の種類に限定することとした。作家間の推薦において重要だと考える関係について、20代の女性18人を被験者としてアンケートを行った結果、最多であり、6割以上の被験者が重要であると回答した“作家の影響関係”が推薦に有効であると判断した。具体的には、DBpedia のオントロジー “<http://dbpedia.org/ontology/influenced>” , “<http://dbpedia.org/ontology/influencedBy>” によってリンクした作家の関連リソースを、着目した作家に対しての推薦候補とする。着目作家から推薦候補の作家まで、影響関係を辿る回数を2者間の「距離」とした場合、DBpedia において作家“村上春樹”を表すリソース^{*3}と、距離が1の作家との関係を図1に示す。

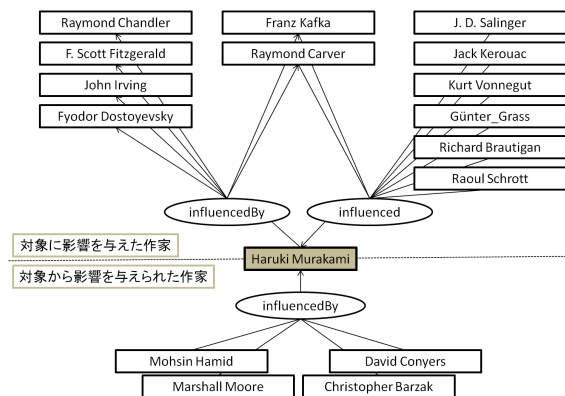


図1: 「村上春樹」から距離が1の関連作家

2.2 リンク解析による推薦情報の発見

得られた推薦候補の評価には、大西ら [1] によって定義されている、RDF で記述されたリソースの性質を測る3つの指標を用いることとした。

注目リソース R と関係によってリンクした要素の集合を $\Omega = \{r_1, r_2, \dots, r_\alpha\}$, その内、R 以外とのリンクを持つ要素の集合を $\Psi = \{r_1, r_2, \dots, r_\beta\} (\Psi \subseteq \Omega)$ とする。このとき、リソース R の Authority Score , Resource Score は

$$AuthorityScore(R) = \alpha$$

$$ResourceScore(R) = \beta$$

^{*3}http://dbpedia.org/resource/Haruki_Murakami

と定義される。また、 Ψ の各要素の Authority Score の中央値を M ，標準偏差を SD とするとき，Authority Score が $M \pm 1SD$ の範囲内である要素の集合を $\Phi = \{r_1, r_2, \dots, r_\gamma\} (\gamma \leq \beta, \Phi \subseteq \Psi)$ とすると，Hub Score は以下のように定義される。

$$HubScore(R) = \sum_{r \in \Phi} \frac{AuthorityScore(r)}{ResourceScore(r)}$$

Φ の設定は Ψ の要素の中で，Authority Score が極端に大きいリソースを除くために行う。

Hub Score は注目リソース R の周知度を表すため，今回の作家の推薦においても，注目作家に関連する推薦候補のうち，周知である作家を測ることができると考えられる。

3 実験

作家“村上春樹”から距離が1，距離が2の作家について Hub Score を求め，上位10名を推薦対象とした場合に，どのような推薦が行えるかについての実験を行った。それぞれの場合の結果を，表1および表2に示す。また，距離が1における関連作家，距離が2における関連作家上位10件と，着目作家“村上春樹”との影響関係を図2に示す。

表 1: 距離が1における関連作家の Hub Score(評価対象 16 件)

Rank	Person	Hub Score	Authority Score	Resource Score
1	Fyodor Dostoyevsky	571.33	310.0	212.0
2	Franz Kafka	423.05	286.0	182.0
3	Kurt Vonnegut	360.74	256.0	153.0
4	Jack Kerouac	345.44	215.0	137.0
5	Raymond Chandler	248.46	202.0	111.0
6	F. Scott Fitzgerald	201.52	178.0	87.0
7	J. D. Salinger	198.58	188.0	88.0
8	Raymond Carver	181.29	171.0	80.0
9	Günter Grass	176.04	154.0	67.0
10	Richard Brautigan	151.40	133.0	64.0

表 2: 距離が2における関連作家の Hub Score(評価対象 374 件)

Rank	Person	Hub Score	Authority Score	Resource Score
1	Stephen King	1079.81	506.0	405.0
2	H. P. Lovecraft	657.18	362.0	260.0
3	Ernest Hemingway	613.64	333.0	224.0
4	Jorge Luis Borges	521.75	307.0	208.0
5	Leo Torstoy	513.08	275.0	186.0
6	Philip K. Dick	490.71	269.0	166.0
7	Ray Bradbury	427.66	257.0	171.0
8	Johann Wolfgang von Goethe	419.26	269.0	165.0
9	Mark Twain	411.53	294.0	162.0
10	Vladimir Nabokov	411.10	290.0	171.0

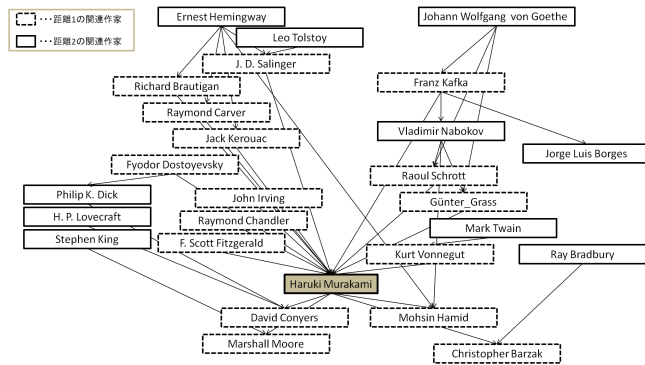


図 2: 距離が2の作家との影響関係

3.1 考察

表1, 2より，“Fyodor Dostoyevsky”，“Stephen King”等，一般的に良く知られた作家の Hub Score が高い傾向にあることが分かる。したがって対象作家のリソースと関係した著名な人物を推薦する場合，Hub Score は有効であると考えられる。また図2より，距離が2の著名な作家を推薦する場合，注目対象との関係の在り方によって，いくつかの異なった推薦の可能性が存在することが分かった。“Stephen King”は表2において，最も評価の高い作家であり，注目作家“Haruki Murakami”とは互いに“Marshall Moore に影響を与えた”という関係を持つ。このとき Haruki Murakami と Stephen King との間に直接の影響関係はないが，互いに同じ人物に影響を与えたという観点で，両者を比較した推薦が行えると考えられる。また，“Ernest Hemingway”は，間接的に Haruki Murakami と5つの関係を持ち，その内4つが“Haruki Murakami に影響を与えた”という関係になっている。この場合，Haruki Murakami は Hemingway から多大な影響を受けたことが推測され，注目作家の作風の源流となった人物の推薦が可能だと考えられる。

4 おわりに

本研究においては，DBpedia 内のデータのリンク解析を行うことによってリソースの関係を調べ，情報を推薦する手法を作家の推薦を例に示した。今後は実験によって明らかになった，距離が2以上の場合における推薦の可能性について検討していきたい。また，今回用いた影響関係以外の関係についても情報推薦の手掛かりとして適用したい。さらに，評価方法を検討することにより，関係の種類に依存しない汎用的な推薦手法を考えていきたい。

参考文献

- [1] 大西可奈子，小林一郎，岩爪道昭，Linked Data を利用した対象文章の情報拡張への取り組み，3E3-OS20-3，人工知能学会全国大会，2011。