

Web 調べ学習における関連概念推薦手法の提案

Recommendation Method for Related Concepts in Web-based Investigative Learning

太田 光一, ティン クン, 長谷川 忍

Koichi OTA, TING Kang, Shinobu HASEGAWA

北陸先端科学技術大学院大学

Japan Advanced Institute of Science and Technology

Email: ota@jaist.ac.jp

あらまし : Web 調べ学習では, 学習者が主体的に Web リソースを選択しながら学習を進めていくため, 設定した学習課題において学習すべき項目やその学ぶ順序(学習シナリオ)を学習者自身で決定しつつ学習を行う必要がある. このとき, ある概念についてそれを取り巻く他の概念と関連付けて認識・記憶しながら学習することでより高い学習効果が期待できるが, 学習中に学習者自身が自らの知識構築を振り返り, 関連する概念を見つけるのは容易ではない. このような問題に対して, 本研究では Linked Open Data(LOD) と PageRank アルゴリズムを用いて, 学習者の主体的な学習を妨げることなく知識構築プロセスに関連する概念(意味的關係)を推薦する手法を提案する.

キーワード : LOD, Web 調べ学習, PageRank, 学習シナリオ

1. はじめに

Web 調べ学習では, 学習項目やその学習順序(学習シナリオ)が予め示されているテキスト教材とは異なり, 学習者自身が非構造的なリソースである Web リソースを選択し, 自ら学習シナリオを作成しながら学習を行っていく必要がある. 先行研究では Web 調べ学習モデルが提案され, このモデルに沿った学習環境として interactive Learning Scenario Builder (iLSB) が開発されている⁽¹⁾.

このとき, ある概念についてそれを取り巻く他の概念と関連付けて認識・記憶しながら学習することでより高い学習効果が期待できる⁽²⁾. しかし学習者自身が, これまでの知識構築をふまえ, 知識構築に関連するような他の概念を自ら探索・収集することは難しい.

このような問題に対して, 本研究では Linked Open Data(LOD) と PageRank アルゴリズムを用いて, 学習者の主体的な学習を妨げることなく知識構築プロセスに関連する概念(意味的關係)を推薦する手法を提案する.

2. Web 調べ学習モデル

本研究では, Web 調べ学習プロセスを以下の 3 フェイズで構成されると定義し, これらのフェイズを新たに部分課題が展開されなくなるまで繰り返すことで, 学習シナリオが作成されることと捉えている.

1. Web リソース探索フェイズ

学習課題を端的に表すキーワードを用い学習に用いる Web リソース群を探索・収集

2. Navigational Learning フェイズ

Web リソースを実際に関連しながら, 学んだ項目からキーワードの抽出・関連づけを行い, 課題に対して学んだ知識を構築

3. 学習シナリオ作成フェイズ

さらに学習が必要な項目を学習課題の部分課題として展開し, 学習課題を構造化

3. 提案手法

本研究では, 1. Web リソース探索フェイズにおいて関連概念を推薦し, 2. Navigational Learning フェイズにおいて自らの知識構築に用いるような学習を想定している.

関連概念推薦にあたり, 次の 3 要素を考慮し推薦を行う.

1. 概念間の意味的關係の生成
2. 抽出された概念の重要度算出
3. 概念の有用性に基づくフィルタリング

これら 3 要素に基づいた推薦手法の詳細を次節以降で述べる.

3.1 Linked Open Data に基づく概念マップの生成

1. 概念間の意味的關係生成にあたり, Linked Open Data (LOD)に着目した.

LOD とは Web 上でデータ同士を結び付けて公開する仕組みのことである. 本研究では, Wikipedia から情報を抽出した LOD である DBpedia を用いる.

DBpedia は RDF と呼ばれる形式のデータ RDF データから構成されるものである. RDF データは主語, 述語, 目的語の 3 つの要素からなる組(トリプル)でリソースに関する関係情報を表現され, クエリ言語 SPARQL を用いることで skos:concept といった概念情報, skos:broader, skos:narrower といった概念の階層関係も検索可能である⁽³⁾. 例えば, <A> broader という関係は, A が B より広い概念を持っていることを意味し, ある“A”に対しての“B”を SPARQL を用いて検索することも可能である.

本研究では、まず知識構築に使われているキーワード群から推薦対象となるキーワードを設定する。次に、キーワードを基に SPARQL を用いて DBpedia から skos:concept を持った RDF データを抽出する。そしてそれらのデータに skos:broader, skos:narrower といった階層構造を返す。

これらの情報をもとに、推薦のベースとなる概念マップの生成を行う。

3.2 PageRank アルゴリズムに基づく概念の重要度算出

PageRank アルゴリズムは、Google が Web ページの重要度を評価するために使用するアルゴリズムとして有名であり、Lawrence Page らによって以下のように定義されている⁽⁴⁾。

$$PR(A) = (1 - d) + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right)$$

PR(A)はWebページAのPageRank(重要度)を表し、PR(T_n)はページAにリンクしているページT_nのPageRank、C(T_n)はページT_nのリンクの数、dはダンピングファクターを表し、0から1の間で設定される。

3.1 で生成された概念マップは相互にリンクしたノードの集合であるため、本研究ではノード(=概念)の重要度は、その概念に関するアウトバウンドリンクの数で決まると仮定し、以下のように定義した。

- ・概念マップの全ノードの数を |R|
- ・リンク {x,r} ∈ E を持つノードxの集合 r ∈ R を B_r
- ・ノードxからのリンクの数を c_x

最終的には、以下の式に基づいて、概念マップのノードrのPageRank値をPR_rとして算出する。

$$PR_r = \frac{1 - d}{|R|} + d \sum_{x \in B_r} \frac{PR_x}{c_x}$$

H. H. Fu らの研究⁽⁵⁾より、d=0.85 と設定した。本研究では、このアルゴリズムを Semantic-aware PageRank とし、3.1 で生成した概念マップに対して Semantic-aware PageRank アルゴリズムを適用し概念の重要度を算出する。

3.3 有用性に基づく推薦提示時のフィルタリング

学習者への関連概念推薦にあたり、DBpedia 上で定義されていない、つまり definition を持たない概念は推薦を行う上で有用性が低いと仮定し、推薦提示時にフィルタリングを行う。

例えば、Natural Language Processing という概念に対して表1の結果が得られた場合、SPARQL で definition を要素に持つ概念を検索することで、Language-specific Linux distributions という概念が DBpedia に定義されていないことを容易に見え、結果としてフィルタリングされることになる。

表1 「Natural Language Processing」における関連概念の推薦候補群

From	Concept	PageRank value
parent	Artificial intelligence applications	0.04195245480414097
parent	Natural language and computing	0.04156089875522573
siblingparent	Character encoding	0.034612167740547484
siblingparent	Computing by natural language	0.031707051327605124
siblingparent	Computational linguistics	0.03020212447579182
siblingparent	Language software	0.028732706859287364
grandparent	Linguistics	0.024832147077640577
siblingparent	Internationalization and localization	0.023271620521276036
child	Corpus linguistics	0.02227424983847515
siblingparent	Language-specific Linux distributions	0.02149842439107931

4. ケーススタディ

理工系大学院生20名を対象として、本提案手法のケーススタディを行った。このケーススタディでは、「機械学習」「原子力」「ガバナンス」「喫煙」というキーワードに対して本提案手法に基づき関連概念推薦リストを作成し、被験者に対してリストを提示することで Web 調べ学習時の推薦システムの有効性を検証した。

これらのキーワードに対し、推薦リストを提示することで、推薦リスト内のキーワードが知識構築時に有効に使われる可能性が示唆された。

5. まとめ

本研究では、Linked Open Data(LOD) と PageRank アルゴリズムを用いて、学習者の主体的な学習を妨げることなく知識構築プロセスに関連する概念(意味的關係)を推薦する手法を提案した。実際に、本提案手法で生成された関連概念推薦リストをもとに、ケーススタディを行った結果、推薦リストが学習者の主体的な学習を妨げることなく知識構築プロセスに寄与する可能性が示唆された。

今後の課題として、関連概念の傾向を踏まえた提示手法の検討、および支援システムの実装とその評価実験が挙げられる。

参考文献

- (1) Akihiro Kashiara, and Naoto Akiyama: "Learning Scenario Creation for Promoting Investigative Learning on the Web", Journal of information and systems in education, Vol.15, No.1, pp.62-72 (2017)
- (2) G. Siemens, "Connectivism: A learning theory for the digital age," International Journal of Instructional Technology and Distance Learning, 2(1), 3-10, 2005.
- (3) 使う・つなげる国立国会図書館の Linked Open Data: <https://www.ndl.go.jp/jp/dlib/standards/loa/index.html>
- (4) Brin, S. and Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. In: Seventh International World-Wide Web Conference (WWW 1998), April 14-18, 1998, Brisbane, Australia.
- (5) H. H. Fu, K. J. Lin, and H. T. Tsai. Damping factor in google pageranking. APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY, 22:431-444, 2005.