

難易度推定アルゴリズムを用いた学習コンテンツ推薦システムの開発に向けた事前調査

Preliminary Research for the Development of Recommendation System Using Algorithm to Estimate the Difficulty of Learning Contents

濱田 一伸^{*1}, 三好 康夫^{*2}, 鈴木 一弘^{*2}, 塩田 研一^{*2}, 岡本 竜^{*2}
Kazunobu HAMADA^{*1}, Yasuo MIYOSHI^{*2}, Kazuhiro SUZUKI^{*2}, Ken-ichi SHIOTA^{*2}, Ryo OKAMOTO^{*2}

^{*1}高知大学大学院総合人間自然科学研究科理学専攻
^{*1} Graduate School of Humanities and Social Sciences, Kochi University

^{*2}高知大学理学部
^{*2} Faculty of Science, Kochi University
Email: khamada@is.kochi-u.ac.jp

あらまし:我々はこれまで学習コンテンツ推薦システムの構築を目指し,コンテンツ難易度推定アルゴリズムの研究を進めてきた.本アルゴリズムは誰が何を読んだかという関係からなる2部ネットワークを分析することによりユーザ習熟度とコンテンツ難易度を推定する.本稿では推薦システム開発に向けて行った推薦対象の検討や先行研究の調査について報告する.

キーワード:コンテンツ難易度推定,ユーザ習熟度推定,学習コンテンツ推薦

1. はじめに

ラーニングなどの学習支援システムにおいて,ユーザの能力に適した難しさの学習コンテンツを推薦する手法は,これまで多く提案されてきた.しかし学術書籍やWebページを学習コンテンツとして推薦対象とする場合,閉じた学習支援システム内のコンテンツのように難易度が自明ではないため,適切な難易度のコンテンツを推薦することは難しい.そこで我々は,学術書籍やWebページを対象に,ユーザ習熟度とコンテンツ難易度を推定するアルゴリズムの設計を行い,学習コンテンツ推薦システムの構築を目指してきた.本稿では,推薦システム構築に向けて調査した内容について報告する.

2. 習熟度と難易度の推定アルゴリズム

2.1 読者ネットワーク

我々が提案するアルゴリズムは,コンテンツの内容に基づいて推定するのではなく,図1に示すような読者ネットワークの情報のみを利用する⁽¹⁾.この読者ネットワークは誰がどのコンテンツを読んだかという関係を表した2部ネットワークで,リンクには読んだ順序を示す番号が付与されている.

2.2 習熟度と難易度の推定の流れ

習熟度と難易度の推定は,以下の手順で行う.

- (1)「ある分野のコンテンツを多く読んでいるユーザはその分野に詳しいはず」と予想し,読んだコンテンツの数に応じた習熟度の初期値を設定する.
- (2)「詳しくないユーザが読んでいるコンテンツは易しいはず」「学習を始めたばかりの頃に読んだコンテンツは易しいはず」などの予想に基づき,ユーザ習熟度からコンテンツ難易度を推定する.
- (3)「難しいコンテンツを読んでいるユーザほど詳しく

いはず」などの予想に基づき,コンテンツ難易度からユーザ習熟度を推定する.

- (4) 手順(2)(3)を収束するまで繰り返し計算する.

3. 推薦システムの開発に向けた事前調査

3.1 既存の情報推薦のオープンソースの調査

我々は,図2のような2段のフィルタリングによる推薦システムを想定している.そこで,オープンソースの協調フィルタリングを調査したところ,Cicindela (<http://labs.edge.jp/cicindela>)とHadoopのMahout (<http://mahout.apache.org>)が見つかった.推薦システムの開発においては,利用が簡単そうなCicindelaを利用する予定である.

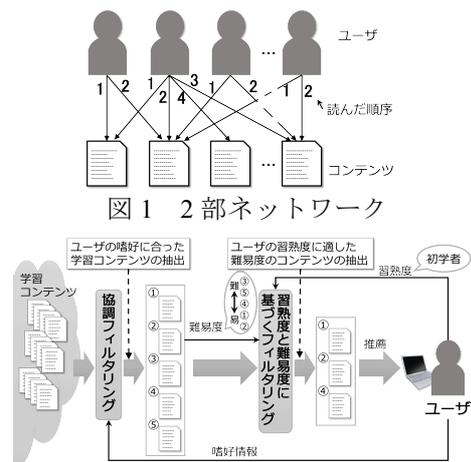


図1 2部ネットワーク

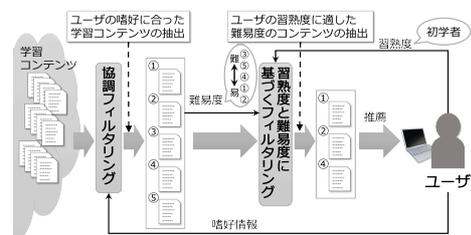


図2 2層フィルタリングに基づく推薦システム

3.2 コンテンツ難易度を考慮した推薦の先行研究

推薦システムを設計するにあたり,難易度を考慮した情報推薦に関する研究が他に行われていないか調査したところ,以下の3つの事例が見つかった.

1つ目は Kille ら⁽²⁾による推薦システムにおける難易度の概念の提案に関する研究である。Kille らが定義する難易度は、ユーザにコンテンツの難しさを評価させた情報に基づいている。これと比較して我々の研究は、ユーザによる評価を必要とせずに難易度を推定できる点で優れていると言える。

2つ目は岩下ら⁽³⁾による協調フィルタリングを利用し英語学習者に適したリスニング教材を推薦する研究である。岩下らの手法もユーザがそれぞれのコンテンツに難しさを評価した点数を登録している。岩下らによると難易度に基づいた推薦結果の適合率はやや低かったとのことであった。その原因は、ユーザの習熟度の違いを考慮せずに、ユーザが付与した点数をそのまま用いたことにあると考えられる。

3つ目は Skocir ら⁽⁴⁾による携帯向けゲームの推薦システムに関する研究である。Skocir らはゲームをプレイしているユーザのスキル(習熟度)からそのゲームの難易度を推定しており、我々のアプローチと類似している。しかしユーザ習熟度の推定においてはゲームのスコアなどを利用できるため、我々の研究と比較して容易に推定可能である。

3.3 推薦対象の検討

学習コンテンツを推薦するにあたり、その推薦対象となるものは、読書履歴に相当するデータを基に読者ネットワークが形成可能なものに限られる。そこで、多くのユーザとコンテンツの情報が登録されていて、読書履歴に相当するデータが収集可能と考えられる対象として、読書管理サイトの読書履歴、図書館の貸出履歴、ソーシャルブックマークサイトのブックマーク履歴の3つを検討した。

読書管理サイトとは、読み終えた本の登録などの機能を持つ Web サービスである。主要な読書管理サイトにはブックログ (<http://booklog.jp>) などがある。しかし主要な読書管理サイトでは、あるコンテンツを登録しているユーザ数は取得できるものの、ユーザ一覧を取得することができない(本稿執筆時)ため、読者ネットワークを作成することができなかった。

図書館の貸出履歴は、COPAC (英国大学図書館総合目録) の SALT recommender API (http://copac.ac.uk/innovations/activity-data/?page_id=227) でマンチェスター大学ジョンライランズ図書館の貸出履歴が公開されている。この貸出履歴には、2001~2011年における、借りた人 191,539 人、書籍 627,999 冊による 3,070,369 件の貸出記録が登録されている。また、ソーシャルブックマークサイトのブックマーク履歴としては、独自にプログラムを作成することで、はてなブックマークからブックマーク履歴の収集が可能となっている。

そこで、図書館貸出履歴とブックマーク履歴からそれぞれ作成した読者ネットワークの比較を行った。比較用の貸出履歴の読者ネットワークは、書籍のタイトルに“PHP”が含まれている本 12 冊、それらの本を借りた 222 人、全体の貸出数 397 件の部分ネッ

トワークである。一方、ブックマーク履歴の読者ネットワークは、はてなブックマークのタグに“PHP”が付与されたブックマークを収集して形成した、Web ページ数 72,962 ページ、ユーザ 15,914 人、ブックマーク総数 315,638 件のネットワークである。

図 3 は、ユーザが PHP に関するコンテンツをいくつか読んでいるかを示したものである。2 章で述べたアルゴリズムではユーザが読んだコンテンツの順序を利用しており、多くのコンテンツを読んだユーザ数が少ないと精度が期待できない。ソーシャルブックマークの読者ネットワークは 3 ページ以上ブックマークしているユーザ数が全体の 62.7% であるのに対し、図書館貸出履歴の読者ネットワークでは 3 冊以上を借りているユーザは全体の 8.56% しかいない。

従って、現時点で推薦システムを実装する際には対象をソーシャルブックマークにブックマークされた Web ページとすることが望ましいと思われる。

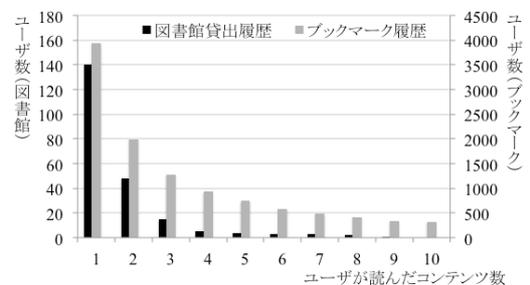


図 3 読んだコンテンツ数ごとのユーザ数の分布

4. おわりに

学習コンテンツ推薦システムに利用可能な協調フィルタリングと、コンテンツ難易度に基づく推薦の事例について調査した。また、推薦対象としては、我々が提案する習熟度と難易度の推定アルゴリズムの特徴に合うソーシャルブックマークの Web ページが適していることを確認した。今後は調査した結果を踏まえ、推薦システムの設計開発を進めたい。

謝辞

本研究は、JSPS 科研費の助成 25330364 を受けた。

参考文献

- (1) 濱田一伸, 三好康夫, 鈴木一弘, 塩田研一: “2 部ネットワーク分析によるユーザ習熟度とアイテム難易度の推定アルゴリズムの提案”, 教育システム情報学会第 37 回全国大会, A6-1, No.2, pp.384-385 (2012)
- (2) Kille, B. and Albayrak, S.: “Modeling Difficulty in Recommender Systems,” WS on Recommendation Utility Evaluation, Beyond RMSE (RUE2011), pp.30-32 (2012)
- (3) 岩下文香, 来住伸子: “協調フィルタリングを利用した英語教材推薦アルゴリズムの研究”, 情報処理学会研究報告 情報学基礎研究会報 2007(34), pp.53-60 (2007)
- (4) Skocir, P., Marusic, L., Marusic, M. and Petric, A.: “The MARS – A Multi-Agent Recommendation System for Games on Mobile Phones,” Agent and Multi-Agent Systems, Technologies and Applications, Springer Berlin Heidelberg, pp.104-113 (2012)