

# 学習コンテンツ推薦に向けた2部ネットワーク分析に基づく 習熟度と難易度の推定手法の提案

三好 康夫<sup>†</sup>, 濱田 一伸<sup>‡</sup>, 鈴木 一弘<sup>†</sup>  
塩田 研一<sup>†</sup>, 岡本 竜<sup>†</sup>, 金西 計英<sup>\*</sup>

<sup>†</sup>高知大学理学部, <sup>‡</sup>高知大学大学院総合人間自然科学研究科

<sup>\*</sup>徳島大学大学開放実践センター

miyoshi@is.kochi-u.ac.jp

**概要** 学習コンテンツを対象とした情報推薦では、ユーザの習熟度に適した難易度のコンテンツが推薦されることが望ましい。そこで、ユーザの習熟度と学習コンテンツの難易度を、誰が何を読んだかという関係を示した2部ネットワーク(読者ネットワーク)のみを用いて推定するアルゴリズムを実装している。読んだ順序を考慮して推定アルゴリズムの精度向上を試み、テストデータを用いて推定精度を評価した。精度評価を行ったことにより、新たに検討すべき課題が明らかになった。

**キーワード** 学習コンテンツ推薦, 習熟度, 難易度, 読者ネットワーク

## 1 はじめに

現在の情報推薦は主にユーザの嗜好情報に基づいて行われている。しかし、参考書や専門書、技術書といった学術書籍や、同様に学術的な内容が書かれたWeb ページ等、学習コンテンツを対象にした情報推薦の場合は、ユーザ(学習者)の習熟度に適した難易度のコンテンツが推薦されることが望ましい。そこで本研究では、ユーザの習熟度と学習コンテンツの難易度のマッチングを考慮に入れたコンテンツ推薦手法の開発を目指している。我々は現在、ユーザの習熟度と学習コンテンツの難易度を推定するアルゴリズムの実装に取り組んでおり、本稿ではその習熟度と難易度の推定手法について述べる。

## 2 読者ネットワーク

我々が提案する習熟度と難易度の推定手法は、図1に示すような読者ネットワークのみを使用する。読者ネットワークは、ユーザとコンテンツ間の、誰がどのコンテンツを読んだかという関係を示した2部ネットワークである。当初は読んだ時間情報を全く使用していなかったが[1]、現在実装中のアルゴリズムではユーザがコンテンツを読んだ順序を考慮している[2]。

アルゴリズムの精度を評価するため、はてなブックマーク (<http://b.hatena.ne.jp>) の“recommendation”タグが付与されたブックマークとそれをブックマークしたユーザの関係からなる読者ネットワークをテストデータとして用意した。ユーザ数は370、ブックマークされたコンテ

ツ数は944、ブックマーク総数は2022である。この読者ネットワークの一部のデータを抜き出し、ユーザの習熟度とコンテンツの難易度の正解データを人手で作成した。正解データ作成手順については文献[3]で述べているため省略する。

## 3 習熟度と難易度の推定手法

### 3.1 基本的な考え

文献[1]で提案したアルゴリズムの詳細は省略するが、基本的な考えは以下の仮説に基づいていた。

- (1) ある分野のコンテンツを多く読んでいるユーザは、その分野について詳しい
- (2) 詳しいユーザに読まれているコンテンツ程難しく、詳しくないユーザに読まれているコンテンツ程易しい
- (3) 難しいコンテンツを読んでいるユーザは詳しく、易しいコンテンツを読んでいるユーザは詳しくない

(2)はユーザの習熟度からコンテンツの難易度を求め、(3)はコンテンツの難易度からユーザの習熟度を求めている。この(2)と(3)を収束するまで繰り返すことで最終的なユーザの習熟度とコンテンツの難易度が求まる。(1)ははじめにユーザの習熟度の見当をつけている。

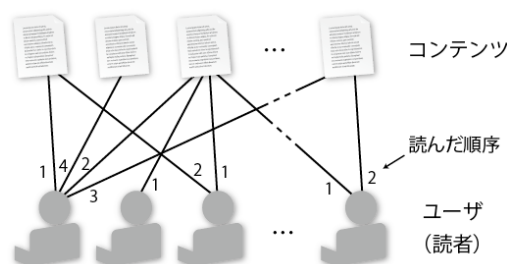


図1 読者ネットワーク

### 3.2 読んだ順序を考慮した推定手法

上で述べた旧アルゴリズムは、中山ら[4]にも追試されたが、あまり高い精度が得られていない。また、パラメータを調整すると、仮説(2)の「詳しいユーザに読まれているコンテンツ程難しい」を反映させないようにしたときが最も良い結果が得られた。確かに詳しい人であっても易しいコンテンツを読むことは大いにありうるし、詳しい人が詳しくなかった頃に読んだのかもしれない。そこで読んだ順序を考慮した推定手法を提案した。

まず、新たな推定手法では習熟度と難易度の定義を見直した。旧手法では習熟度と難易度の互いの値は直接的に関係がなく、ある習熟度のユーザに適したコンテンツの難易度の値を一意に決められないため、情報推薦の際の基準としては扱いにくかった。そこで、図2のような知識獲得量モデルを考え、それに基づいて次のように定義した。ユーザ  $u$  の習熟度  $P_u$  は、対象2部ネットワークの分野に関して  $u$  が持つ知識量とする。コンテンツ  $i$  の難易度  $D_i$  は、それを読むのに必要な知識量とする。ユーザ  $u$  は  $P_u = D_i$  のときに最も多く知識をコンテンツ  $i$  から得られる。

旧手法ではユーザの習熟度  $P_u$  を求める際、ユーザが読んだコンテンツ全体から最終的な習熟度を1つ推定していたが、新手法ではコンテンツを読むごとに獲得した知識分の習熟度を増加させる。つまり、図2のようにコンテンツを3つ読んだ場合、初期の習熟度  $P_u(0)$  を含めて4つの習熟度  $P_u(0) \sim P_u(3)$  を推定する。そしてコンテンツの難易度  $D_i$  は、ユーザの最終的な習熟度ではなく、そのコンテンツ  $i$  を読み始めたときの習熟度を基に推定する。なお、初期習熟度  $P_u(0)$  は図2では  $D_{i1}$  より低くなっているが、実装中のアルゴリズムでは  $P_u(0) = D_{i1}$  としている。ただし、次に読んだコンテンツの難易度の方が最初に読んだコンテンツより易しい場合は、 $P_u(0)$  を次に読んだコンテンツの難易度に下げる。

## 4 推定精度向上に向けての考察

新手法を実装したアルゴリズムの動作確認を2章で述べたテストデータで行った。その結果を表1に示す。精度を評価するための基準にはYaoが提案したNDPM

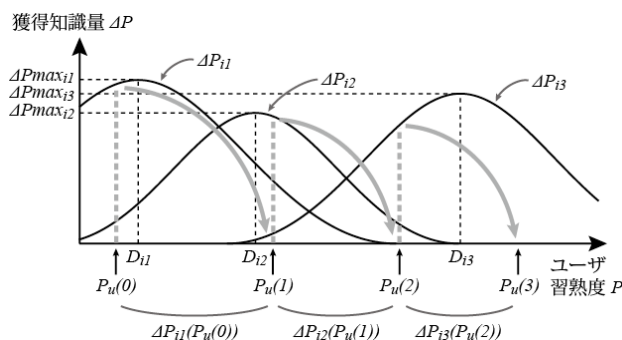


図2 知識獲得量モデル

表1 難易度の求め方毎の習熟度と難易度のNDPM値

| 難易度の求め方                 | NDPM 値 |        |
|-------------------------|--------|--------|
|                         | 習熟度    | 難易度    |
| 旧アルゴリズム <sup>[1]</sup>  | 0.2887 | 0.4598 |
| 読んだユーザの習熟度の平均値          | 0.3557 | 0.3423 |
| 読んだユーザの習熟度の中央値          | 0.3763 | 0.3135 |
| 読んだユーザの習熟度の平均値と最小値との中間値 | 0.2732 | 0.3501 |
| 読んだユーザの習熟度の中央値と最小値との中間値 | 0.2990 | 0.3135 |
| 読んだユーザの習熟度のうち2番目に小さい習熟度 | 0.2320 | 0.3266 |
| 読んだユーザの習熟度の最小値          | 0.2732 | 0.4259 |

値を用いた[5]。NDPM 値は小さいほど、アルゴリズムが求めた習熟度や難易度の順位と正解データの順位との差異が小さく、精度が良いことを示す。コンテンツ  $i$  の難易度  $D_i$  を求める際、 $i$  を読み始める各ユーザの習熟度の平均値あるいは中央値を  $D_i$  とすれば良いと考えていたが、表1に示すように、実際にはそれより低いところを  $D_i$  とした方が良い結果が得られていた。従って、 $i$  を読むユーザの多くは  $D_i$  より高い習熟度なのではないかと推測され、 $\Delta P_i$  の鐘形曲線は再検討する必要があると考えられる。今後は、項目応答理論[6]等を参考に、分析を進めてさらなる推定精度の向上を目指したい。

## 謝辞

本研究の一部は、平成22年度文部科学省科研費挑戦的萌芽研究(課題番号:22650203)と若手研究(B)(課題番号:22700815)の補助を受けた。

## 参考文献

- [1] 三好康夫, 入野美弥: 学術書籍の難易度を読者ネットワークから推定する試み, 電子情報通信学会教育工学研究会技術研究報告, Vol.110, No.67, ET2010-5, pp.19-24, 2010.
- [2] 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36回全国大会講演論文集, pp.384-385, 2012.
- [3] 濱田一伸, 三好康夫, 金西計英: アイテム難易度とユーザ習熟度を推測するアルゴリズムを評価するための正解データ作成手法, 日本教育工学学会研究報告集, JSET12-2, pp.113-118, 2012.
- [4] 中山祐輝, 南保英孝, 木村春彦: レビュー情報を用いた学術本の難易度推定, 人工知能学会論文誌, Vol.27, No.3, pp.213-222, 2012.
- [5] Yao, Y. Y.: Measuring retrieval effectiveness based on user preference of documents, Journal of the American Society for Information Science, Vol.46, No.2, pp.133-145, 1995.
- [6] 大友賢二: 項目応答理論入門, 大修館書店, 1996.