

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

Estimation Algorithm of User Proficiency and Item Difficulty with Bipartite Network Analysis

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

^{*1}高知大学大学院総合人間自然科学研究科理学専攻

^{*2}高知大学理学部

^{*1} Graduate School of Humanities and Social Sciences, Kochi University

^{*2} Faculty of Science, Kochi University

Email: khamada@is.kochi-u.ac.jp

あらまし：情報推薦において、推薦アイテムが学習コンテンツである場合、ユーザの習熟度に合った難易度のアイテムが推薦されることが望ましい。本研究では、誰がどのアイテム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アイテム難易度を推定するアルゴリズムを提案する。提案アルゴリズムは、ユーザがアイテムをどの順序で読んだかを考慮した仮説に基づく反復計算により習熟度と難易度を求める。

キーワード：習熟度、難易度、学習コンテンツ、2部ネットワーク

1. はじめに

既存の情報推薦技術はユーザの嗜好情報に基づいて推薦を行うが、推薦アイテムが学習コンテンツである場合、ユーザの習熟度に合った難易度のアイテムが推薦されるべきである。そのため、嗜好情報だけではなく、ユーザ習熟度とアイテム難易度を考慮した推薦が必要と考えられる。そこで本研究では、誰がどのアイテムを読んだかという関係からなる2部ネットワーク（読者ネットワーク）の情報のみを用いてアイテム難易度とユーザ習熟度を推定するアルゴリズムを提案する。我々は先行研究において、読者ネットワークからユーザ習熟度とアイテム難易度を推定するアルゴリズムの試作を試みた⁽¹⁾。本論文ではこれを従来アルゴリズムとし、問題点の解消を目指した改良アルゴリズムの提案を行う。

2. 従来アルゴリズム

2.1 仮説

従来アルゴリズムの基本的な考え方は、以下の仮説に基づいている。

- (1) ある分野のアイテム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アイテムほど難しく、詳しくないユーザに読まれているアイテムほど易しい。
- (3) 難しいアイテムを読んでいるユーザは詳しく、易しいアイテムを読んでいるユーザは詳しくない。

この仮説のうち、(2)と(3)が相互に影響し合い、(2)ではユーザ習熟度からアイテム難易度を求め、(3)ではアイテム難易度からユーザ習熟度を求めている。

(2)と(3)を繰り返し計算することでユーザ習熟度とアイテム難易度が求まる。(1)でははじめにユーザ習熟度の見当をつけている。

2.2 計算の流れ

2.1節の仮説を式に表すと式1)、2)のようになる。

$$a_u = \sum_{i \in I_u} d_i \dots \dots \dots 1)$$

$$d_i = \sum_{u \in U_i} (a_u - 0.5) \dots \dots \dots 2)$$

u : ユーザ a_u : 習熟度 \hat{a}_u : 正規化した習熟度

i : アイテム d_i : 難易度 \hat{d}_i : 正規化した難易度

I_u : u が読んだアイテム U_i : i を読んだユーザ

最初に、習熟度 a_u の計算を行う。初期値は、習熟度 a_u は 0、難易度 d_i は 1.0 とする。仮説(1)よりユーザが読んだアイテムの数だけ習熟度 a_u に難易度を加算し、最後に全てのユーザの習熟度を正規化(偏差値 ÷ 100)することで習熟度が求まる。難易度 d_i は、仮説(2)より読まれた各ユーザの習熟度を加算し、正規化することで求まる。これらの計算を習熟度と難易度が収束するまで繰り返す。

2.3 問題点

詳しいユーザでも易しいアイテムを読むことがあるということを仮説(2)では考慮しておらず、仮説として不十分である。これにより従来アルゴリズムの推定精度は期待できない。実際に中山ら⁽²⁾による評価実験の結果においても、我々の従来アルゴリズムが推定したアイテム難易度の精度は低かった。

また従来アルゴリズムでは、習熟度と難易度の値は順位尺度としての意味しか持たず、互いの値には直接的な関係はない。つまり、習熟度 a_u のユーザに適切なアイテムの難易度がどのような値をとるか

いうことは不明である。したがって、従来アルゴリズムで推定した習熟度と難易度は、情報推薦の際の基準として扱いにくい。

3. 読んだ順序を考慮した改良アルゴリズム

本章では 2.3 節で述べた従来アルゴリズムの問題点の改善を目的とし、ユーザのアイテムを読んだ順序を考慮して習熟度と難易度を推定する改良アルゴリズムについて述べる。

3.1 習熟度と難易度の定義

アイテムを読んだ順序を考慮するため、2 部ネットワークは図 1 のようにユーザと読んだアイテム間のリンクには読んだ順序を示す番号が付与される。

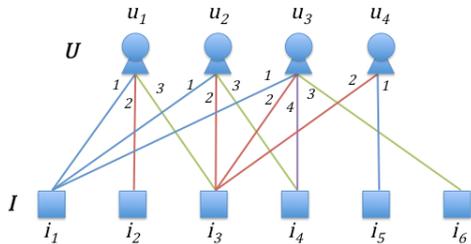


図 1 読んだ順序を含めた 2 部ネットワーク

改良アルゴリズムの開発にあたり、ユーザ習熟度とアイテム難易度を定義しておく。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。また、アイテム i の難易度 D_i は、そのアイテムを読むのに必要な知識量を意味し、 $P_u = D_i$ のときにユーザ u がアイテム i から最も多くの知識を得ることができるものと定義する。これにより、情報推薦の際には、習熟度 P のユーザに対して難易度 D が P に近いアイテムを推薦すればよく、2.3 節の 2 つ目の問題点を解決することができる。

3.2 読んだ順序を考慮して習熟度と難易度を計算する流れ

改良アルゴリズムも従来と同様に、アイテムの難易度からユーザの習熟度を求め、またユーザ習熟度からアイテム難易度を求め、これを収束するまで交互に繰り返す。改良アルゴリズムでは、それぞれを求める際に、読んだ順序を考慮し、ユーザがどのように知識を獲得していったかを推測することで精度向上を目指す。図 2 を用いてその方法を説明する。

ユーザ u がアイテム i_1, i_2, i_3 の順に読んだとき、ユーザ習熟度 P_u は初期習熟度 $P_u(0)$ から $P_u(3)$ へと知識が獲得されていく。 $P_u(t)$ は t 番目のアイテムを読んだ後の習熟度である。ある 1 つのアイテム i を読んだときに得られる知識量を ΔP_i とすると、同じアイテムを読んだとしても、 ΔP_i はそのときのユーザ u の習熟度 P_u によって変化する。例えば図 2 において、 u がアイテム i_1 を読んで獲得する知識量 $\Delta P_{i1}(P_u(0))$ は、 $P_u(0)$ が D_{i1} より少し低いため、最大獲得知識量 $\Delta P_{max_{i1}}$ より少し少ない量となる。そして $P_u(1) = P_u(0) + \Delta P_{i1}(P_u(0))$ となる。

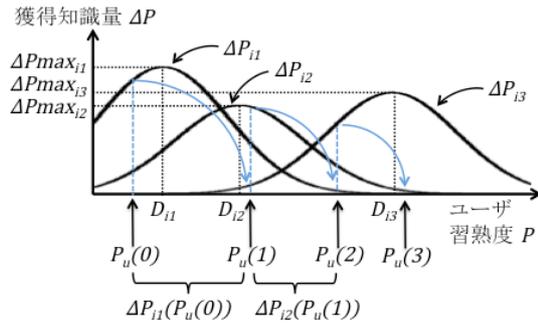


図 2 ユーザ u の知識獲得量

なお、初期習熟度 $P_u(0)$ は、初めに読んだアイテムの難易度とする。ただし、次に読んだアイテムの難易度の方が初めに読んだアイテムより易しい場合は、 $P_u(0)$ は次のアイテムの難易度に下げる。よって、図 2 では D_{i1} より $P_u(0)$ が低くなっているが、試作中の実際のアルゴリズム上では、 $P_u(0) = D_{i1}$ となる。

ΔP_i の鐘形曲線を形成する式は、 D_i と ΔP_{max_i} と曲線の幅を決める σ_i からなる。アイテム i の内容が充実しているほど ΔP_{max_i} は大きくなり、 i を読んだユーザたちの習熟度の幅が広いほど曲線の幅が広がるよう σ_i は変化するものと考えられるが、その算出方法と鐘形曲線式については紙面の都合上省略し、ここでは難易度 D_i の算出方法についてのみ述べる。

難易度 D_i は、アイテム i を読むのに必要な知識量を意味していることから、 i を読んでいたユーザたちの大半は読み始める際の習熟度が D_i と同程度か D_i より高いはずである。したがって、アイテム i を読み始める際の各ユーザの習熟度の平均値あるいは中央値より低いところが D_i になる。試作中のアルゴリズムにおいては、 i を読み始める際の各ユーザの習熟度の最小値と中央値との中間値を D_i としている。

4. おわりに

現在、ソーシャルブックマークデータから生成した 2 部ネットワークを用いてアルゴリズムの検証を行っている。今後の課題は、既に作成済みの正解データ⁽³⁾を用いて精度を評価することである。

謝辞

本研究の一部は、平成 22 年度文部科学省科研費若手研究(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) 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)
- (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定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”, 日本教育工学会研究報告集, JSET12-2, pp.113-118 (2012)