

オープン教育リソースのための順序推定

中島 克也[†] 小林 亜樹[†]

[†] 工学院大学

1 はじめに

インターネット上で教材 (オープン教育リソース; OER) を配布し、ユーザの学習を支援する活動が盛んに行なわれている。ユーザが教材を検索したとき、ユーザの選択した教材は既存知識で理解できるものとは限らず、閲覧を始めてからユーザの持ち合わせていない知識が要求されるたびにその都度新たに教材を検索し、選択しなおすのは非常に煩わしい。

そこで、教材ごとに前提となる教材や、次に学ぶべき教材を共に推薦することが可能ならば、ユーザの学習効率の向上につながる。そのためには、教材間の順序情報をも提供することが必要であるが、教材の数の膨大さゆえに、教材の提供者がすべての教材との関係を把握し、順序情報を付与することは事実上不可能である。

Web ページに対しては順序判定を行なう西原らによる研究 [1] があるが、体系化された教育コンテンツに順序判定を行なう例は未だない。そこで本研究では提案手法として、教材間の順序情報の機械判定を行なうことを目的とする。

2 提案手法

2.1 方針

本研究では、教育リソースへ付与された著者説明文およびコンテンツ、タグ集合、コメントなど UGC を含むテキスト集合が得られるとき、これらのテキストを解析し、“学習分野” 及び“前提分野” を得ることで順序判定を行なう。

ここで、将来的には大規模教育リソースへの適用を目指す。分野の表現方法や付随するテキストメタデータの適切性の検討が必要である。また、手法の評価のためには正解順序を得ておく必要もある。そのため本稿では、教育リソースのメタデータによる順序推定の可能性を探るため、正解順序を得やすく、一定の質を持つ大学シラバステキストを用い、分野にはテキスト中の名詞句を利用する基本的なテキスト解析を試みる。

2.2 準備

M 個の教育コンテンツを要素とする集合を $R = \{r_i | 1 \leq i \leq M\}$ とする。 $r_i \in R$ は教育コンテンツを表す。 r_i に付随する“学習内容”を示すテキストを T_i 、“前提内容”を示すテキストを P_i とする。

R のすべての要素 $r_i (1 \leq i \leq M)$ から T_i 及び P_i を取得し、それぞれ 1 文書として考える。総文書数は $2M$ である。関数 noun を、テキスト t を与えたとき、 A に含まれる名詞句を要素とする集合 A へ変換するものと定義する。また、関数 $\text{tfidf}(n)$ を、名詞句 n を与えたとき、 n の tf-idf 値を算出するものと定義する。

2.3 提案手法の手順

$\forall r_i, r_j (i \neq j)$ について、 $N_{ij} = \text{noun}(T_i) \cap \text{noun}(P_j)$ とし、スコアを S_{ij} とする。

$$S_{ij} = \sum_{n \in N_{ij}} \text{tfidf}(n) \quad (1)$$

S_{ij} を要素としてもつ行列 $\mathcal{R} = \{S_{ij}\}$ を計算する。時系列的に前に学習すべきコンテンツを判定する際には i 列目、後に学習すべきコンテンツを判定する際には i 行目からなる要素を、 S について降順に並べ替え、ユーザへ提示することが可能となる。

3 実験

3.1 方針

本稿では、工学院大学のシラバスデータに本手法を適用し、実験を行なう。工学院大学の各講義におけるシラバスデータは、講義の目標を示す〈授業のねらい〉、講義を受講するにあたって事前に習得しておくべき分野や講義を示す〈受講にあたっての前提条件〉、講義の到達目標を示す〈具体的な到達目標〉、各回の講義の内容を示す〈授業計画及び準備学習〉といった項目が記述されている。

そこで、各講義に付随するシラバスデータの〈授業のねらい〉の項に記された文章を“学習分野”、〈受講にあたっての前提条件〉の項に記された文章を“前提分野”とみなすことで本提案手法を適用し、次の実験を行なった。

3.2 条件

工学院大学情報通信工学科の 207 講義において、時系列的に後に学習すべき講義を S の大きいものから順

Learning sequence estimation for open educational resources

[†] Katsuya Nakajima (c510086@ns.kogakuin.ac.jp)

^{††} Aki Kobayashi (aki@cc.kogakuin.ac.jp)

に表示する。取得したテキスト数は414である。

また各講義において S の大きいものから4個を対象にし、F-measureのマクロ平均で評価する。正解順序には、工学院大学から公式に配布されている、講義を時系列順に並べたフローチャートを基に作成したテストコレクションを用いてF-measureを測定し、マクロ平均を求める。

講義 i に対して S の降順に4つ抽出し、テストコレクションにおいて講義 i を前提とする全ての講義と比較し、一致するものを適合文書とする。

また、測定対象とする講義 i において、講義 i を前提とする講義が存在しないものを除く。測定対象の講義数を53とする。

3.3 実験結果

適切な結果が得られたものの一例を以下に示す。事実上順序関係の存在する講義の S_{ij} の値が高いと判断されたものを「適切」であるとみなす。

適切な結果が得られたものの一例を以下に示す。

表 1: 適切な結果の得られた例

順位	数学 II		線形代数学 II	
	講義名	S	講義名	S
1	微分方程式論	1.02	幾何学 II	1.41
2	物理学実験	0.92	応用解析学	1.16
3	幾何学 II	0.90	微分方程式論	1.15

適切な結果が得られなかったものの一例を以下に示す。

表 2: 適切な結果の得られなかった例

順位	芸術学入門		心理学入門	
	講義名	S	講義名	S
1	化学実験	0.37	線形代数学 I	0.57
2	経済学入門	0.27	通信ソフトウェア I	0.48
3	国際経済と金融	0.27	応用解析学	0.45

適切な結果が得られたものは相対的に見て S が大きな値をとるものが多く、適切な結果が得られなかったものは S が相対的に小さな値をとることが見て取れた。

また、評価結果の一例を表3に、F-measure値の平均値を表4に示す。

3.4 考察

名詞句の一致を順序判定基準とし、tf-idf値をスコアとして与えるという提案手法の構造上、付属する文書集合に含まれる単語の種類数が多いものほど S の値が大きくなる傾向にあり、これを要因として適切な順序判定を行なうことのできない事例が見受けられた。

また、F-measureの測定に用いたテストコレクションの作成に大学から提供されるフローチャートを用い

表 3: 評価結果の一例

数学 II	
正解後続科目	確率・統計学 微分方程式論 複素関数論 応用解析学 R.H.I デジタル信号処理 線形システム論 情報通信工学実験 I V 画像工学 マルチメディア工学
精度	0.25
再現率	0.1
F-measure	0.14285
後続科目	S
微分方程式論	1.024763339
物理学実験	0.924374967
幾何学 I I	0.902814608
数学演習 I	0.899497457

表 4: F-measure の平均値

評価対象講義数	53
F-measure の平均値	0.2167

だが、事実上関連性があるであろう講義間においても、関係性が反映されていない例が多数存在したため、適切な学習順序判定が行なわれていると判断できる講義であっても、F-measureの値が非常に低い値をとる場合が多く見られた。テストコレクションの適切な与え方は今後の課題である。

4 おわりに

教育リソースの学習順序自動判定をめざして、説明テキストからの判定手法を提案した。提案手法のうち、必要なテキストの抽出を行うことが可能であるとするときに本手法の有効性を確認するため、大学のシラバスデータによる検証実験を行った。テキスト抽出と組み合わせる実際のデータによる有効性の検証は今後の課題である。

謝辞

本研究の一部は科学技術研究費（基盤（A））No. 26242013による。

参考文献

- [1] 西原 陽子, 砂山 渡, 谷内田 雅彦, “Web ページの難易度と学習順序に基づく情報理解支援システム”, 信学論, J89-D(9), pp.1963-1975, 2006-9