

# 高校科目における単元と大学科目の類似度を用いた授業理解度向上に関する一考察 A Study on the Improvement of Course Comprehension Using the Similarity of Units in High School Courses and University Courses

浮田 善文<sup>†</sup> 齋藤 友彦<sup>††</sup> 松嶋 敏泰<sup>‡‡</sup>

Yoshifumi Ukita Tomohiko Saito Toshiyasu Matsushima

## 1 はじめに

大学での授業において授業理解度の低い学生を考えた場合、理解度が低い原因の一つとして、関連する高校科目の単元の理解不足があげられる。このため、内容の関連が強い高校科目の単元が分かれば、この単元にまで立ち返り、学修することで、授業理解度の向上が期待できる。ただし、大学での開講科目は科目数も多く、また内容も多岐に渡るため、各科目の授業担当者、関連する高校科目の単元を明らかにしてもらうことは、作業負担の面からも困難であった。

ここで大学での開講科目全体を考えると、科目数が多く、科目ごとに情報を集めることは困難なため、公開されているシラバス情報を用いることが考えられる。シラバスを用いた科目間の類似度を利用した研究は既に行われており、履修科目の推薦システムに関する研究 [1, 2] や授業理解度向上に関する研究 [3] などがある。ただし、これまで授業理解度向上に用いられていたのは科目間類似度のみで、科目と単元の類似度を利用するといった研究はこれまで見られなかった。

そこで、本研究では、公開されているシラバス情報により得られる科目と単元間の類似度と、科目関連グラフを用いることで、授業担当者が少ない作業負担で、関連が強い単元を選択可能な方法を提案する。

## 2 シラバス情報による科目間類似度

本章では、本稿で用いるシラバスデータとシラバス情報による科目間類似度について説明する。科目間類似度を測る方法には、TF-IDF 値を利用する方法、Doc2Vec を利用する方法、BERT を利用する方法など、様々な方法が考えられる。次章で説明する科目関連グラフは、科目間類似度を求められるのであれば、どの方法を用いることもできる。以下では一例として、TF-IDF 値を用いた方法について説明する。

### 2.1 シラバスデータ

2022 年度に横浜商科大学で開講した全科目 (複数コマ開講している科目は 1 科目としてカウント) から、ゼミナール科目を除いた 275 科目に対し、科目ごとに

科目情報として、シラバスに記載されている「授業の目的と概要」、「到達目標」、「内容」を用いた。また、高等学校の単元については、実教出版 [4] の教科書 (5 科目: 情報 I, 数学 I, 数学 A, 数学 II, 数学 B) に含まれる「章タイトル」からなる計 21 単元とした。科目情報については、実教出版教科書のダウンロードデータに含まれる年間指導計画案の中の「学習内容や学習活動」と「主な評価基準」を用いた。

以下で、高校の単元は科目としてもみなし、大学 275 科目と高校 21 科目の合計 296 科目としても扱うこととする。

### 2.2 TF-IDF 値を利用した科目間のコサイン類似度

本節では、296 科目に対し、形態素解析エンジンである MeCab [5] を用い、名詞、形容詞、動詞を抽出した。抽出した単語の出現頻度を用い計算される各科目の TF-IDF 値のベクトルを Python 用のライブラリ scikit-learn を用いて求めた。

また、科目間の内容の類似度には、ベクトル空間におけるコサイン類似度を用いる。科目  $i$  と科目  $j$  の TF-IDF 値のベクトルをそれぞれ、 $\mathbf{c}_i$ ,  $\mathbf{c}_j$  とすると、コサイン類似度  $cs(\mathbf{c}_i, \mathbf{c}_j)$  は次式により計算される。

$$cs(\mathbf{c}_i, \mathbf{c}_j) = \frac{\mathbf{c}_i \mathbf{c}_j}{\|\mathbf{c}_i\| \|\mathbf{c}_j\|} \quad (1)$$

以下で例として、大学科目「情報社会の倫理」を取り上げ、高校科目の単元との関連を見ていくことにする。高校科目の 21 単元のうち、大学科目「情報社会の倫理」とコサイン類似度が高い上位の 12 単元を表 1 に載せる。

表 1: 大学科目「情報社会の倫理」との類似度

順位	科目名	類似度
1	情報 I: 情報社会	0.2652
2	情報 I: 情報デザイン	0.1308
3	情報 I: ネットワーク	0.0927
4	情報 I: デジタル	0.088
5	情報 I: プログラミング	0.0814
6	数学 I: データの分析	0.0788
7	情報 I: 問題解決	0.0771
8	数学 A: 数学と人間の活動	0.0738
9	数学 I: 数と式	0.071
10	数学 II: 複素数と方程式	0.0647
11	数学 A: 図形の性質	0.0636
12	数学 B: 数学と社会生活	0.0625

表 1 より、想定していた通り、類似度の 1 位は「情報 I: 情報社会」であった。一方、関連は少ないと考え

<sup>†</sup> 横浜商科大学 商学部 経営情報学科  
Yokohama College of Commerce

<sup>††</sup> 湘南工科大学 情報学部 情報学科  
Shonan Institute of Technology

<sup>‡‡</sup> 早稲田大学 理工学術院 Waseda University

られる「数学 II: 複素数と方程式」と「数学 A: 図形の性質」が 10 番目と 11 番目に来ているが、これらの科目の類似度が低いことから、類似度の値も重要であることが分かる。

次に、高校科目の単元を取り上げ、大学科目との関連を見ていくこととする。一例として、高校情報 I の単元「情報社会」とコサイン類似度が高い大学科目の上位 12 科目を表 2 に載せる。

表 2: 高校情報 I の単元「情報社会」との類似度

順位	科目名	類似度
1	知的財産権法	0.3247
2	情報社会論	0.2661
3	情報社会の倫理	0.2652
4	経営情報学 2	0.2471
5	ビッグデータ解析	0.2385
6	心理学	0.2045
7	財務諸表分析	0.2026
8	商取引法	0.2007
9	英文会計	0.2004
10	ICTリテラシー 1	0.1962
11	ネットワークとコミュニケーション 2	0.195
12	財務諸表論	0.1903

表 2 を見ると、「情報社会論」と「情報社会の倫理」が上位に来るのは予想通りである。一方、科目名からは予想していなかった「知的財産権法」、「心理学」、「商取引法」といった科目が上位に来ており、科目名だけでは気づきにくいものの類似度が高い科目を見つけられることが分かる。

### 3 科目関連グラフ

この章では、選択した科目に対し、関連の強い科目全体を視覚的に把握するための科目関連グラフを提案する。ここで、科目の関連度をコサイン類似度により測ることとする。

以下、グラフについて説明する。まず、科目および単元をグラフの点とし、この点の集合を  $V$  とする。 $V$  の 2 点  $u, v$  に対し、コサイン類似度が  $t$  以上であれば  $u, v$  間に辺があるとし、この辺の集合を  $E$  とする。このとき、点の集合  $V$  と辺の集合  $E$  からなる図形を、コサイン類似度  $t$  のグラフ  $G_t = (V, E)$  と定義する。

グラフ  $G_t$  の点の有限列  $P = \langle v_0, v_1, \dots, v_n \rangle$  が  $v_{i-1}v_i \in E (1 \leq i \leq n)$  を満たしているとき、 $n$  を  $P$  の長さという。 $v_0, v_n$  はそれぞれ  $P$  の始点、終点と呼ばれる。

科目  $s$  に対して、 $s$  を始点とする長さ  $n$  以下の全ての有限列に含まれる点の集合を  $V_s, (V_s \subset V)$  とする。 $V_s$  の任意の 2 点  $u, v$  に対し、コサイン類似度が  $t$  以上となる辺の集合を  $E_s, (E_s \subset E)$  とする。このとき、グラフ  $G_{t,n}^s = (V_s, E_s)$  を科目  $s$  に対する類似度  $t$ 、長さ  $n$  の科目関連グラフと呼ぶ。

一例として、科目関連度として TF-IDF 値によるコサイン類似度を用いた場合について、JavaScript によ

り作成した、科目「情報社会の倫理」に対する類似度 0.21、長さ 2 の科目関連グラフを図 1 に示す。

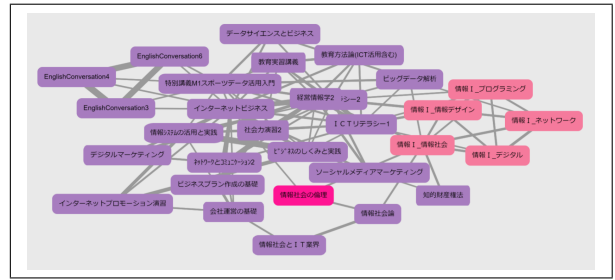


図 1: 科目「情報社会の倫理」の科目関連グラフ

図 1 では、濃いピンク色の点を選択した「情報社会の倫理」である。今回は、大学科目と高校科目の単元の 2 グループに分け、グラフの点の色については大学科目は紫色、高校科目の単元は薄いピンク色とした。今回は 2 グループのみとしたが、さらに、学年単位、学科単位、専門科目群単位などの細かいグループ分けも可能である。グループごとにグラフに表示するかどうかも選択可能にしているため、例えば、高校科目の単元と大学 1 年の科目の関連のみを見ることも可能である。また、辺である線の太さは、コサイン類似度の大きさに対応している。

さらに、選択した科目からの有限列の長さも指定可能である。これにより、選択した科目との二者間での関連度が高い科目のみを見たいのであれば長さを  $n = 1$  とし、関連度の高い科目群がさらにどのような科目群とつながっているかを知りたいのであれば長さ  $n$  を大きくすることで可能である。

### 4 類似度と科目関連グラフを用いた内容の関連が強い高校科目の単元選択

この章では、授業担当者に、少ない作業負担で、担当科目と内容の関連が強い高校科目の単元を選択してもらう方法を提案する。具体的には、以下の通りである。

まず、授業担当者に、担当科目との類似度が高い高校科目の単元のリスト (例: 表 1)、および、科目関連グラフ (例: 図 1) を提供する。授業担当者は、必要に応じて科目関連グラフを参考にしながら、類似度が高い単元のリストの中から、内容の関連が強い単元を選択する。

ここで選択された単元を、シラバスなどで受講の前提となる知識に加えることで、受講開始前に、学生が単元の重要性を知り、高校での学習内容を復習することが期待できる。また、この単元に対する確認テストで学生の理解度を把握し、理解が不足している学生には、単元の復習課題を課すことで、大学での受講科目

の理解度向上が期待できる。

また、科目関連グラフは、大学で開講している全ての科目間の繋がりを把握することが可能であるため、これまで知られていなかった異分野の科目間の繋がりなどについても知ることができる。

ここまでは、科目間の類似度として、TF-IDF 値による類似度を用いてきたが、他の類似度についても考える。

表 3 に、2 種類の類似度 (TF-IDF 値による類似度と BERT による類似度) について、大学科目「情報社会の倫理」とコサイン類似度の高い高校科目の上位 12 単元を載せる。なお、BERT は最長トークン数を 256 とし、日本語の事前学習モデル\*を利用し、ファインチューニングなしで実施した。

表 3: 類似度の比較 (「情報社会の倫理」との類似度)

順位	TF-IDF 値による類似度	BERT による類似度
1	情報 I:情報社会	情報 I:情報社会
2	情報 I:情報デザイン	数学 I:データの分析
3	情報 I:ネットワーク	情報 I:ネットワーク
4	情報 I:デジタル	情報 I:情報デザイン
5	情報 I:プログラミング	数学 B:数学と社会生活
6	数学 I:データの分析	数学 I:集合と論証
7	情報 I:問題解決	数学 I:数と式
8	数学 A:数学と人間の活動	数学 II:いろいろな関数
9	数学 I:数と式	数学 I:三角比
10	数学 II:複素数と方程式	数学 A:場合の数と確率
11	数学 A:図形の性質	数学 I:2 次関数
12	数学 B:数学と社会生活	数学 A:数学と人間の活動

表 3 より、各類似度の上位 12 単元のうち、両方に含まれる単元は 7 単元であった。このように、利用する類似度により、類似度の上位の単元も変わることが分かる。このため、授業担当者に、複数の方法で得られる類似度上位の単元の情報を全て提供することも考えられる。

## 5 おわりに

本稿では、大学での授業担当者に、担当科目との類似度が高い高校科目の単元のリストと科目関連グラフを提供することにより、少ない作業負担で、関連が強い高校科目の単元を選択してもらう方法を提案した。また、関連の強い単元の情報を利用した、学生の授業理解度向上への取組についても考察を行った。

今回は、実際に選択した単元を利用することの効果までは検証できていないため、この検証が今後の課題としてあげられる。また、学生の単元の理解度を効率よく把握するためのテストの作成、理解度に応じた単元の復習課題の作成も今後の課題である。

## 謝辞

本研究は JSPS 科研費 JP22K02811 の助成を受けたものです。

\* <https://github.com/cl-tohoku/bert-japanese>

## 参考文献

- [1] 西森友省, 堀幸雄, 今井慈郎: 履修履歴を用いた科目推薦システム, 情報処理学会第 75 回全国大会, pp.645-646 (2013).
- [2] 竹森汰智, 亀井清華: 履修支援のための Doc2Vec を用いた科目推薦システム, 情報処理学会論文誌データベース, Vol.12, No.4, pp.1-14 (2019).
- [3] 浮田善文, 齋藤友彦, 松嶋敏泰: シラバス情報による科目間類似度を用いた授業理解度向上に関する一考察, 情報処理学会第 85 回全国大会講演論文集, 4, pp.283-284, (2023).
- [4] 実教出版高等学校教科書シラバス (2023), 入手先 (<https://www.jikkyo.co.jp/highschool/>).
- [5] MeCab, available from (<https://taku910.github.io/mecab/>).
- [6] 近江崇宏, 金田健太郎, 森長誠, 江間見亜利: BERT による自然言語処理入門, オーム社, (2021).