

シラバスを用いた学習院女子大学における科目間ネットワークの分析

清水 将吾・澤田 匡人

Abstract

Gakushuin Women's College offers a wide variety of subjects as a liberal arts college. On the other hand, the wideness makes it difficult for students to grasp the overall picture of the curriculum. In this paper, we classify subjects based on the course number assigned to each subject, and perform network analysis to quantitatively evaluate the density of relationships among subjects and identify subjects that play a central role in each field.

1 はじめに

学習院女子大学国際文化交流学部は教養系の学部として様々な分野の科目を提供している。一方で学習者にとっては科目の選択の幅が広く、カリキュラムの全体像が把握しづらいという側面がある。本稿では、各科目に割り振られているコースナンバーに基づいて科目を系や分野ごとに分類し、それらの科目群に対してネットワーク分析を行うことで、各分野の科目間の関連性の密度や各分野において中心的な役割を果たしている科目を定量的に評価することを目的とする。科目間ネットワークの構築にはシラバスのテキストを利用する。

2 準備

2.1 関連研究

シラバスのテキストを活用して科目間の関連性を評価する研究は多数行われている。最も有名な手法として、MIMA Search^{*1}[4]がある。MIMA Searchではオントロジーを構築して科目間の類似度を評価し、ネットワーク構造を使って可視化することで履修計画の支援を図っている。文献[7]でも科目間の類似性をtf-idfにより評価した同様のシステムを開発し、ユーザ評価を行っている。文献[6]では新入生に教養科目を推薦することを目的とした履修支援システムの開発を行っている。推薦の際には各科目のシラバスを

^{*1} <https://ocw.u-tokyo.ac.jp/search/>

Doc2Vec を用いて表現し、クラスタリングを行って科目群の特徴を可視化している。文献 [1] では Doc2Vec に基づく科目推薦手法の改良を行っている。本稿も科目のベクトル化までは同様の手法を採用するが、その後にネットワーク分析を行う点でこれらと異なる。

文献 [3] では、異なる学科間でのカリキュラム比較を行うために、科目群をどの学科でも同様の概念を扱う科目と学科に特異な科目に分類する手法を提案している。本学ではほとんどの科目で学科を超えての履修が可能であるため、学科に特異な科目を抽出する意義があまりない。文献 [2] では、スライドを対象として、スライドに含まれるテキストを利用して時系列を考慮した tf-idf によりスライド間の関連付けを行う手法を提案している。時系列の考慮は履修順を評価する際に有用であるが、教養系学部では履修順の制約が厳しくないことが多く、本稿では順序の概念は取り入れていない。

学習者間の関係に着目した研究も行われている。文献 [5] では、学生の履修科目から科目間の共起ネットワークを構築し、文理融合系学部の学生が幅広い選択肢の中から履修する科目が、自身が専攻するコースに大きく影響を受けていることを確認した。文献 [8] では、学習者のネットワークが学習効果に与える影響についてシミュレーション分析を行っている。これらの手法は本学の履修状況の把握にも適用可能であるが、シラバスを用いた分析とは出発点が異なる。

2.2 科目および出現語の分布

学習院女子大学の 2022 年度シラバスのデータを、Python Beautiful Soup を用いた Web スクレイピングにより「講義コード」、「授業科目」、「副題」、「コースナンバー」、「授業の到達目標」、「授業の内容及び授業の方法」、「授業計画」（実施週ごと）の各項目ごとに取得した。

各科目に対して「授業の到達目標」、「授業の内容及び授業の方法」、「授業計画」のそれぞれに MeCab を使用して形態素解析を行い、得られた単語の中から名詞のみを抽出した。数字や囲み文字、記号のみからなる文字列は削除した。以降の分析を日本語で行うため、アルファベットのみからなる文字列も削除した。日本語ストップワードとして SlothLib^{*2}を使用した。この時点で日本語の単語数が 20 以下の 215 科目は分析対象から除外し、得られた科目総数は 1,020 となった。

各科目の学問分野はコースナンバーにより体系付けられている^{*3}。まず、系として「総

^{*2} <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>

^{*3} 学習院女子大学学生便覧, <https://www.gwc.gakushuin.ac.jp/faculty/syllabus.html>

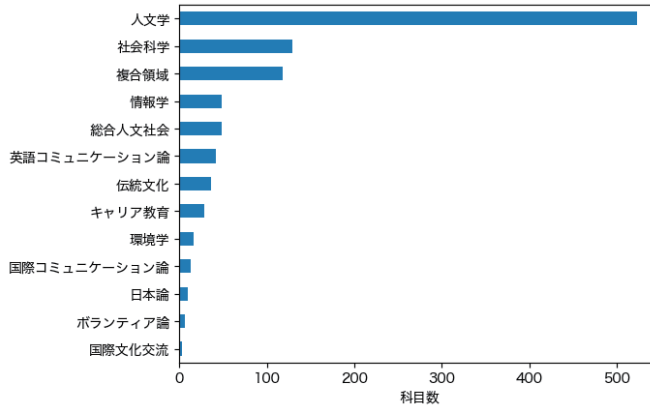


図1 各分野の科目数

合系」「人文社会系」「その他」に分かれており、各系がそれぞれ3, 3, 7の分野に分かれている。「総合系」には182科目、「人文社会系」には700科目、「その他」には138科目が分類されている。各分野の科目数は図1の通りである。

単語の種類は6,850語、出現回数の合計は117,518である。表1に、全体および各系での頻出語上位20件を示す。同一シラバスで複数開講している科目も重複して数えている。また、一つの科目内で同じ単語が繰り返し出現する場合も重複して数えている。したがって、出現頻度の高い語が様々な科目で扱われているとは限らない。

各分野における特徴語を抽出した結果を表2に示す(数値は略)。ここでは、科目数の多い4分野の上位20語のみを示している。特徴語はtf-idfにより評価した。tf-idfとはある語が文書に出現する頻度と、その語が他の文書にどの程度少ない頻度で出現するかという語の特異性を掛け合わせて算出される。つまり、出現頻度が高い語であっても、どの文書にも出現するような一般的な語であれば低く評価される。表2の結果は上記定義における文書を当該分野に出現するすべての語として計算したものである。

2.3 科目データのベクトル化

得られた単語列から科目データをベクトル化する。ベクトル化の手法はword2vec^{*4}を採用する。word2vecはニューラルネットワークを用いて単語の分散表現を得る手法であり、単語を意味空間上のベクトルとして表現することで単語に対して意味的な計算を行

*4 <https://code.google.com/p/word2vec/>

表 1 各系における頻出語

全体	頻度	総合系	頻度	人文社会系	頻度	その他	頻度
発表	2,119	文化	466	発表	1,569	講義	413
授業	1,984	発表	451	授業	1,449	日本	378
日本	1,841	授業	327	日本	1,271	英語	327
研究	1,591	国際	314	研究	1,208	授業	208
文化	1,470	研究	288	理解	908	香	206
理解	1,315	体験	276	語	878	文化	180
社会	1,042	環境	257	文化	824	理解	178
語	934	食	255	社会	752	国際	178
論文	840	情報	252	報告	691	世界	158
講義	832	コミュニケーション	244	論文	617	演習	156
英語	826	理解	229	表現	578	作成	148
国際	782	資料	220	討論	519	表	140
報告	744	スポーツ	217	問題	497	基本	137
問題	714	実践	214	英語	491	計算	130
表現	675	課題	205	学生	488	ソフト	124
学生	644	活動	201	内容	459	使用	118
まとめ	638	問題	199	まとめ	425	実習	115
基礎	605	日本	192	文法	421	指導	113
教育	605	分析	187	テーマ	421	章	113
文献	602	社会	186	練習	408	社会	104

うことができる。日本語の事前学習済みモデルとしては fastText*⁵を使用する。fastText は Web 上のテキストをクロールして得られた Common Crawl*⁶と Wikipedia を基に学習し、単語に対して 300 次元のデータを生成する。例として、同モデルにおいて「機械学習」と類似性の高い語として「乱択アルゴリズム」「組合せ最適化」「ニューラルネットワーク」などが得られ、表記の一致によらず意味的に近い語が抽出できている。

科目のベクトルは、その科目に含まれるすべての単語の word2vec の平均として求める。科目間の類似度はベクトル間の cos で評価する。これにより、例えば、「情報処理 I」と最も近い科目は「情報処理 III」であり、「情報処理 I」と「情報処理 III」の科目間の類似度 0.94 は「情報処理 I」と「教育心理学」の科目間の類似度 0.74 より高いという計算が行える。

*⁵ <https://fasttext.cc/>

*⁶ <https://commoncrawl.org/>

表 2 各分野における特徴語

人文学	社会科学	複合領域	総合人文社会
発表	報告	文化	東欧
日本	研究	スポーツ	アジア
授業	教育	コミュニケーション	研究
研究	授業	体験	文化
文法	社会	発表	討論
会話	発表	研究	個別
読解	理解	国際	講読
論文	学校	ファシリテーション	地域
英語	改革	授業	プレゼンテーション
練習	日本	博物	社会
表現	国際	活動	授業
文化	問題	遺産	論文
理解	指導	食物	東南
動詞	分析	実践	文献
文学	個別	文献	発表
報告	討論	環境	テーマ
作文	企業	教育	報告
言語	学習	人間	イスラム
単語	文献	ならび	移動
学生	チーム	資料	理解

3 科目ネットワークの特徴の分析

3.1 科目ネットワークの概要

科目を節点とし、科目ベクトル間の類似度が 0.9 以上 1 未満の 2 科目に対して辺を定義する。最大連結成分の節点数は 807、辺の数は 9,154 であり、以降この最大連結成分のみを対象とする。図 2 に構築された科目間ネットワークを示す。図中、節点の色は系を示しており、赤は「総合系」、青は「人文社会系」、黄は「その他」の科目である。最大連結成分の次数分布は図 3 の通りであり、平均次数は 22.7 である。半数弱の科目が 10 以下の科目と関連しており、概ねベキ法則に従っているとみなせる。

図 4 に「人文学」、「社会科学」、「複合領域」分野の科目ネットワークを示す。

3.2 分野別のネットワーク指標

分野別にリンク密度、平均最短経路長、クラスタ係数を求める。

リンク密度は可能なすべての辺の数に対する実際の辺の数の割合であり、次式で定義さ

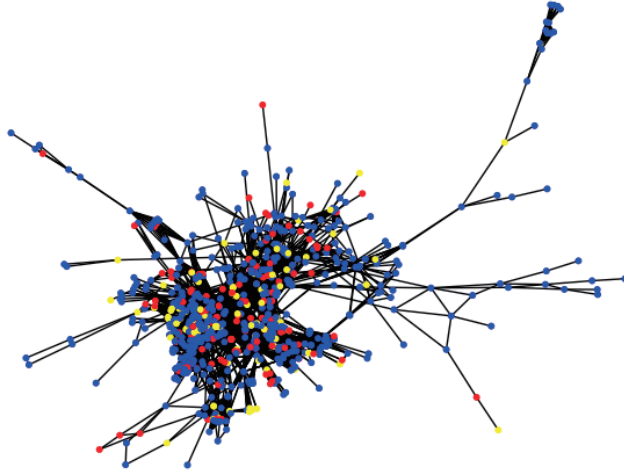


図2 科目間のネットワーク

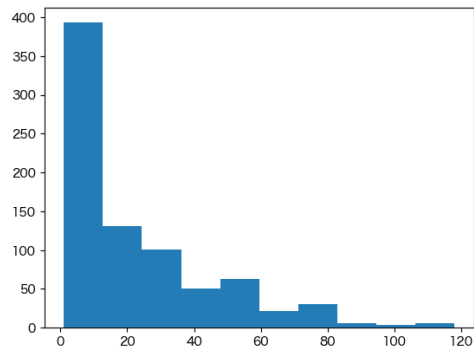


図3 次数分布

れる。

$$\frac{m}{n(n-1)/2}$$

n は節点の数, m は実際の辺の数である。

最短経路長とは, ある節点から別の節点に到達する経路に含まれる辺の数の最小値である。平均最短経路長はすべての節点の組合せに対する最短経路長の平均である。ここで

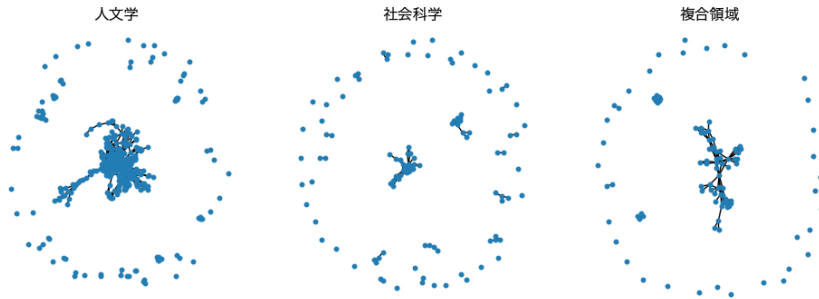


図4 分野ごとの科目ネットワーク

は、各分野の最大連結成分に対して平均最短経路長を求める。

ある節点 i のクラスタ係数 C_i は、その隣接節点同士に辺が存在する割合として定義される。科目 A が科目 B と科目 C に関連があるとき、科目 B と科目 C にも関連がある割合が高ければ、クラスタ性が高いと言う。ネットワークのクラスタ係数はすべての節点のクラスタ係数の平均である。定義は以下の通りである。

$$C_i = \frac{E_i}{\frac{k_i(k_i-1)}{2}}$$

$$C = \frac{1}{N} \sum_i C_i$$

ここで、 k_i は節点 i の次数、 E_i は i の隣接節点間に存在する辺の数、 N は節点の数である。リンク密度がネットワーク全体の密度を表すのに対し、クラスタ係数は局所的な密度を表している。

結果を表3に示す。リンク密度は全体では0.022であり、どの分野も関連科目の割合に大きな差はない。全体の平均最短経路長は4.14であり、任意の2科目に対して、間に平均で3科目程度介せば関連性がたどれることになる。科目数の多い「人文学」と「社会科学」を比較すると、リンク密度は同程度であるが、平均最短経路長とクラスタ係数は「人文学」の方が大きいことから、「人文学」の方が局所的な範囲で密につながっている、「社会科学」の方が広範囲に緩やかにつながっていることがわかる。この結果は、直感的には「人文学」には多くの各国語の科目が含まれており、それらの科目間の局所的な関連性の強さが影響していると考えられる。

表3 各分野のネットワーク指標

分野	リンク密度	平均最短経路長	クラスタ係数
情報学	0.026	3.83	0.248
環境学	0.064	3.83	0.231
複合領域	0.025	3.78	0.278
総合人文社会	0.020	3.78	0.138
人文学	0.020	4.44	0.465
社会科学	0.023	2.26	0.282
伝統文化	0.048	1.94	0.300
キャリア教育	0.017	1.94	0.000
ボランティア論	0.000	0.00	0.000
日本論	0.000	0.00	0.000
国際コミュニケーション論	0.000	0.00	0.000
英語コミュニケーション論	0.121	1.79	0.471
国際文化交流	0.000	0.00	0.000
全体	0.022	4.14	0.550

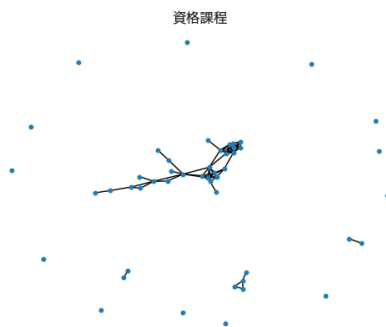


図5 資格課程の科目ネットワーク

3.3 資格課程のネットワーク

コースナンバーとは別に、教職課程、学芸員課程、司書課程の計 56 科目の枠組みでネットワークの特徴を調べる。資格課程における関連の数を十分に確保するため、ここでは類似度の閾値を 0.85 以上に設定した。結果のネットワークは図5の通りである。

類似度の閾値 0.85 での各ネットワーク指標の比較を表4に示す(抜粋)。コースナンバーでの分類と比較して関連の存在する割合が低く、局所的な密を示すクラスター係数も低くなっている。これは、資格課程が少ない科目数で重複なく広範な学習項目をカバーできるよう設計されていることが影響していると考えられる。逆に、それ以外の科目群は緩

表 4 資格課程のネットワーク指標

分野	リンク密度	平均最短経路長	クラスタ係数
複合領域	0.121	2.33	0.530
人文学	0.103	2.49	0.632
社会科学	0.105	2.30	0.536
資格課程	0.057	3.10	0.319

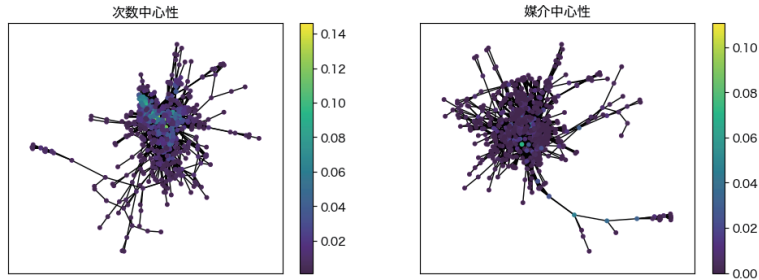


図 6 科目の中心性

やかに冗長性があり、学習状況に応じて様々な選択が可能とみなすことができる。

4 科目ネットワークの中心性の分析

科目の中心性を図 6 に示す。次数中心性はネットワークの中央に指標の高い節点が集中しているのに対し、媒介中心性はネットワークの右下にも指標の高い節点が見られる。これらの節点は右下の科目群と中央の科目群をつなげるために必ず経由する科目になっている。

科目の中心性を次数中心性、媒介中心性により評価する。

節点 i の次数中心性 dc_i は、 k_i を i の次数、 n を節点の総数としたとき、次式で定義される。

$$dc_i = \frac{k_i}{n-1}$$

「人文学」「社会科学」「複合領域」の次数中心性を表すネットワークを図 7 に示す。

節点 i の媒介中心性 bc_i は、任意の 2 つの節点の最短経路に i が含まれる確率として定義される。

$$bc_i = \sum_{s,t \in V} \frac{\sigma(s,t|i)}{\sigma(s,t)}$$

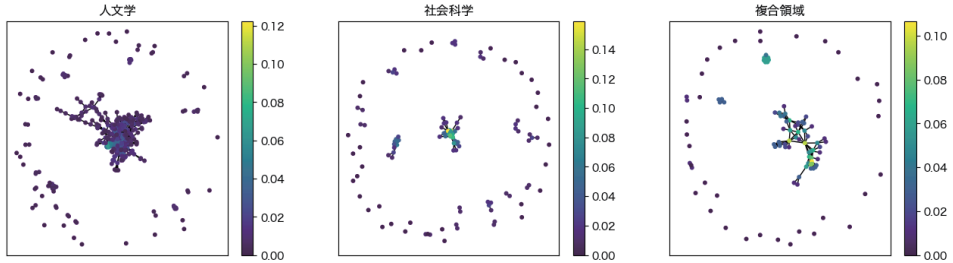


図7 分野ごとの次数中心性

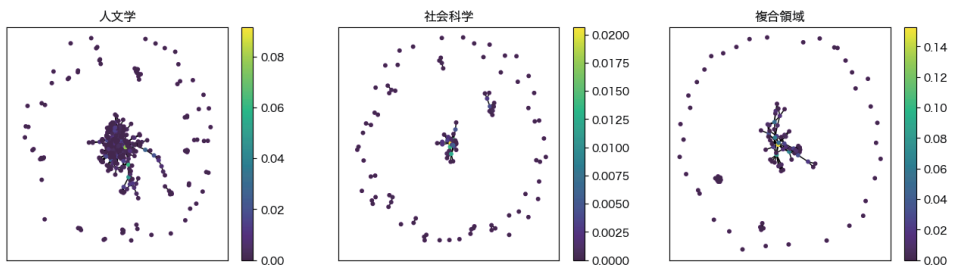


図8 分野ごとの媒介中心性

ここで、 V は節点の集合、 $\sigma(s, t)$ は s, t 間の最短経路の数、 $\sigma(s, t | i)$ は s, t 間の最短経路のうち、途中に i が含まれている経路の数である。「人文学」「社会科学」「複合領域」の媒介中心性を表すネットワークを図8に示す(数値は略)。

次数中心性では、他の科目との関連が多いハブ的な役割を果たしている科目が高く評価される。媒介中心性では、他の科目との橋渡しになるような科目が高く評価される。「人文学」「社会科学」「複合領域」分野のそれぞれの上位10科目を表5に示す。特に、「複合領域」における「国際関係基礎論I」は媒介中心性が0.1であり、全体の中で2番目に高い「国際コミュニケーション演習IIU」の0.07と比較しても顕著な媒介性を示している。これらのうち最も媒介中心性が低い分野は「社会科学」であり、どの関連性のつながりにも含まれるような中心的な科目が少なく、領域に応じて科目が分かれていることがわかる。一方で、「人文学」では「日本語表現法」のような履修モデル上の必修選択科目が実際に媒介中心性の上位に出現していることが確認できる。また、教職課程では「教育相談」「国語科教育法III」、学芸員課程では「博物館資料保存論」など、資格課程科目において中心的な役割を果たしている科目も散見される。

表5 各分野における中心性の高い科目

人文学	回数中心性	媒介中心性
	中国語 基礎 I C	イタリア語 基礎 I B
	中国語 基礎 II C	日本語表現法 I L
	国際コミュニケーション演習 IV V	READING & WRITING L
	国際コミュニケーション演習 III V	西洋芸術論
	日本文化演習 II D	BASIC LISTENING B
	中国語 基礎 II B	日本語表現法 II F
	Writing Practice B	日本文化演習 II D
	フランス語 基礎 I B	中国語 基礎 I C
	BASIC LISTENING N	中国語 基礎 II C
	TOEIC Skills A	INTENSIVE READING & WRITING D
社会科学	回数中心性	媒介中心性
	教育相談	教育相談
	現代文化論 II (郊外論)	国際コミュニケーション基礎演習 IV J
	国際コミュニケーション基礎演習 IV J	日本社会論
	国語科教育法 III	現代文化論 II (郊外論)
	国際文化交流論 I (文化と感情)	国語科教育法 III
	国際コミュニケーション演習 II E	学校インターンシップ
	国際コミュニケーション演習 I M	国際コミュニケーション演習 II C
	国際コミュニケーション演習 I E	国際文化交流論 I (文化と感情)
	国際コミュニケーション演習 II C	国際法 I
	国際法 I	国際コミュニケーション演習 I M
複合領域	回数中心性	媒介中心性
	国際コミュニケーション演習 II U	国際関係基礎論 I
	国際関係基礎論 I	国際コミュニケーション演習 II U
	比較文化論 VI (嗜好)	日本文化演習 IIN
	国際コミュニケーション演習 III A	日本文化基礎演習 IV L
	国際コミュニケーション基礎演習 IV Q	日本文化基礎演習 I K
	国際文化交流論 IX (国際機関)	比較文化論 VI (嗜好)
	日本文化基礎演習 I L	国際コミュニケーション演習 III A
	生活文化演習 I (染織)	国際コミュニケーション基礎演習 IV Q
	博物館資料保存論	国際コミュニケーション基礎演習 I A
	スポーツ・健康科学演習 I A	日本文化演習 I L

5 おわりに

本稿では、学習院女子大学のシラバスを対象として、科目間ネットワークの分析を行った。今回はすべての科目を対象に俯瞰的な視点で分析を行ったが、今後は特定の科目群に着目してより詳細な分析を行う予定である。科目のベクトル化には「発表」や「討論」な

ど授業の方法に関する語も使用したが、分野の専門用語と切り分けて作成する手法も検討する。また、他カリキュラムとの比較や履修支援への適用方法も今後の課題とする。

謝辞

本研究は学習院女子大学令和4年度特別研究費の助成を受けた。

参考文献

- [1] 橋本菜々子, 越智洋司, “シラバスデータを活用した用語からの科目推薦手法の検討”, 教育システム情報学会 2022 年度第 2 回研究会, pp.26–29, 2022 年 7 月.
- [2] 印部太智, 丸山一貴, “自習中の学生のための科目を超えた関連ページの発見”, 情報処理学会研究報告コンピュータと教育 (CE), pp.1–8, 2022 年 3 月.
- [3] 熊田大雅, 佐藤哲司, “学科の特異性を明らかにするための科目概念の推定手法”, 第 14 回データ工学と情報マネジメントに関するフォーラム (DEIM2022), F34–1, 2022 年 3 月.
- [4] Hideki Mima, “MIMA Search: A Structuring Knowledge System towards Innovation for Engineering Education”, Proc. of the COLING/ACL 2006, pp.21–24, 2006
- [5] 塩澤大輝, 松澤芳昭, “文理融合系学部における履修モデル可視化システムの開発と評価”, 情報教育シンポジウム (SSS2019), pp.206–213, 2019 年 8 月.
- [6] 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 年 10 月.
- [7] 山本雄介, 峰松翼, 長沼祥太郎, 谷口雄太, 大久保文哉, 島田敬士, “科目の関連性情報を付加したカリキュラム情報閲覧システムの開発”, 情報処理学会研究報告コンピュータと教育 (CE), pp.1–8, 2021 年 12 月.
- [8] 安武公一, 山川修, 多川孝央, 隅谷孝洋, 井上 仁, “ネットワーク・コミュニティを通じた学習者間の相互作用とその効果に関するシミュレーション分析”, 教育システム情報学会誌, Vol.28, No.1, pp.50–60, 2011.

(本学准教授, 本学准教授)