

自然言語処理を用いた情報工学に関連する 講義の不足単元可視化に関する研究

松崎 由幸[†] 中村 嘉志[‡]

国土館大学 理工学部理工学科 電子情報学系[†]

1. はじめに

近年、社会のデジタル・トランスフォーメーション（以降、DX）により大学教育でも令和3年より文部科学省が認定した「数理・データサイエンス・AI教育プログラミング認定制度」（以降、認定制度）^[1]がある。本学でも本年度より、認定制度によりデータサイエンスに関する講義が開講された。認定制度では、DX促進の大きなきっかけの1つである人工知能（以降、AI）を活用することができる人材を増やすことを目的としており、情報工学を学ぶことはAIを学術的観点から学ぶ上で最低限必要である。

本学の理工学部電子情報学系でも電気・電子工学から情報工学まで幅広く学ぶことが可能なカリキュラムが用意されている。カリキュラムにて幅広い学問分野に対応していることは学際的な知識の獲得ができる利点がある。しかし、自身の関心がある特定の分野でどれほど不足しているのかを学生が判別することは困難である。例えば、本学の電子情報学系選択科目には情報工学の基礎であるブール代数を扱う講義が明確な形では開講されていない。

この問題を解決する方法として、講義シラバスを外部の基準と照らし合わせて不足単元を提示することが挙げられる。比較において様々な基準が考えられるが、比較の対象として本学の講義シラバスに出現する情報工学に関連する専門用語に着目する。

本研究では、講義シラバスに記載されている専門用語を外部の基準と照らし合わせ、不足単元の可視化を目的とする。これにより、情報工学で自身の関心のある分野を学びたい学生は講義で補えていない単元の把握が可能になる。講義シラバスから単元を取得するための科目対象として、令和2年度理工学部便覧に記載されている情報系専門科目とする。外部の基準には、学問の対象を情報工学全般としているため、情報工学の国際標準を基に情報処理学会が編纂した「カリキュラム標準 J-17」^[2]（以降、J-17）を基準として対照に用いる。

ここで、講義シラバスを基に J-17 と単純に比較し、不足単元を導出することは困難である。なぜなら、講義シラバスでは“オペレーティングシステム”や“AI”と J-17 では“OS”や“人工知能”と専門用語の表記ゆれが発生している。また、各講義が対応している分野について、内容を抽象的に記載しているため、情報工学の基礎単元に関する情報が J-17 とは粒度が異なる。

そのため、記載されている情報の粒度を統一する必要がある。

そこで本研究では、不足単元の抽出の際に講義シラバスと J-17 に対して、自然言語処理を用いて講義シラバスと J-17 に記載されている専門用語の表記ゆれを統一し、単元を比較する。また、比較結果を基に不足単元の可視化できるかを検証する。

2. 関連研究

関谷ら^[3]は、カリキュラム改善のために大学の計算機科学関連カリキュラムを外部の基準と比較した。外部の基準には、計算機科学関連カリキュラムである「J07-CS」を用いた。J07-CS との比較の際に、関谷らは記載内容の詳細度を統一するため、大学のシラバスと情報処理学会の作成した教科書を用いた。比較結果に基づき、それぞれ生成したグラフを重ね合わせることでマップを生成し、マップからカリキュラムの分析および改善ができることを示した。

本研究では関谷らが行った手法を踏襲して分析するが、次の2点において異なる。1つ目は、カリキュラム標準 J-07（以降、J-07）の代わりに J-17 を比較基準に用いる点である。J-17 は、J-07 で定められた6個のカリキュラム標準に加えてサイバーセキュリティに関するカリキュラムである J17-CyberSecurity が追加されている違いがある。

2つ目は、語彙の表記ゆれに対して Latent Dirichlet Allocation（以降、LDA）を施す前に辞書を用いて統一する点である。LDA とは、確率的モデルに基づいて文書の集合から個々の文書の特徴づける潜在的なトピックを抽出する手法である。また、LDA による解析結果の出力においては、Python の標準ライブラリである pyLDAvis を用いる。

3. 提案手法

本研究では講義シラバスおよび J-17 に記載してある専門用語の表記ゆれを統一する。頻出単語の抽出後講義シラバスと J-17 を比較する。提案手法の概要を図1に示す。

3.1 前処理

本学情報系専門科目のシラバスを参照するために、国土館大学 Web シラバスから対象科目をテキストデータ化し、J-17 が定めた領域ごとにまとめる。次に両者に共通して出現する表記ゆれした専門用語に対して辞書^[4]を用いて統一する。その際に表記ゆれした専門用語の一例を表1に示す。その後、両者に形態素解析を行い、一般名詞を対象として講義ごとに頻出単語を抽出する。

Visualization of missing units in lectures related to information technology using natural language processing

[†]Yoshiyuki Matsuzaki [‡]Yoshiyuki Nakamura

[†]Faculty of Science and Engineering, Department of Science and Engineering, Faculty of Electronics and Information Science

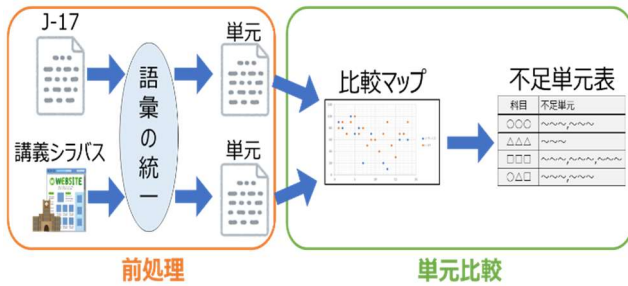


図1 提案手法の概要

表1 表記ゆれした専門用語の一例

カテゴリ	専門用語
省略語	OS
外来語	オペレーティングシステム
和訳語	デジタル
	ディジタル
直訳不可語	AI
	人工知能
	HCI
	ヒューマンコンピュータインタラクション
集約語	ブル代数
	状態遷移図, 真理値表etc.
技術用語から学術用語への変換語	経営戦略
	システム戦略

3.2 単元比較

J-17では、単元があらかじめ領域ごとに定義されている。しかし、本研究ではそれらの単元をLDAの確率的モデルとして再定義し、比較対象に含まれる専門用語群を1つの文書と見なす。これにより、J-17の各トピックがどの程度本学で扱う単元に関連しているかを求められる。

LDAにて抽出してきた各トピックについての関連度をグラフにする。これを用いて、グラフよりJ-17に対応した領域を定め、講義シラバスの内容と重ね合わせることで不足単元を可視化したマップを生成する。このマップを比較の指標に用いて、本学の情報工学の講義から学ぶことが可能なJ-17が定めた単元に適合しているか、また何が不足している単元なのかを判別する。

4. 結果および考察

生成したマップを基に比較し、不足単元の抽出およびマップを作成した。J-17と本学の専門科目講義の比較により、J-17の単元に対して講義シラバスの単元が不足している割合が多い結果を示した。この原因として、専門科目には情報工学以外の学問も含まれていることが挙げられる。一例として、画像音響処理が挙げられる。これは、J-17で単元にデジタル信号処理が含まれているが、この講義は情報工学では無く画像・音響工学の学問領域である。その他にも、情報工学と経営学が含まれる情報戦略や、通信工学と情報工学が含まれる情報理論など、情報工学以外の学問も含んでいたため本研究で使用したJ-17には関連度が低かったと考えられる。

続いて、提案手法による外部の情報工学に関する講義シラバスでも不足単元の抽出が可能であるかを検証した。本研究では、ある大学の情報学部情報学科のシ

表2 単元比較の一致率

専門領域	本学		ある大学		
	一致率	不足率	一致率	不足率	
J-17	Cyber Security	42%	58%	76%	24%
	CE	38%	62%	78%	22%
	IT	50%	50%	94%	6%
	SE	13%	87%	86%	14%
	IS	30%	70%	81%	19%

ラバスを用いた。J-17と本学講義単元やJ-17とある大学の比較結果から、講義シラバスの単元の総数に対してJ-17に一致した単元の割合を専門領域ごとに数値化したもの表2に示す。

検証結果として、大学ごとの情報工学の講義についての特徴を捉えた抽出が可能であると考えられる。本学のシラバスでは「プロセス」や「プログラミング」などに関連する単元を補った授業が充実しており、情報系の実習が不足していることが分かった。一方、ある大学の単元一致率が全領域で高い要因として、シラバスがJ-17に記載されている単元よりも詳細かつ豊富に記載している。そのため、J-17のグラフでは重複してした単元がある大学のシラバスでは分散し、より精度の高い単元の抽出が可能になる。これらの結果より、本研究の提案手法では各大学シラバスの不足単元を抽出すること、また特徴を捉えることが可能であると考えられる。

5. まとめ

本研究では、講義シラバスに記載されている単元をJ-17と比較し、自然言語処理を用いて不足単元を可視化した。講義シラバスを基にJ-17と単純に比較するには、両者の情報工学の基礎単元に関する情報の詳細度が課題である。しかし、両者に記載されている専門用語に着目し表記ゆれを統一することで、本学での不足単元が可視化でき、本学シラバスの不足分野の特徴が分かった。また、他大学のカリキュラムを使用しても同じ結果を得られることから不足単元抽出に関する妥当性があると考えられる。

今後の課題として、本研究では情報工学を対称に不足単元を抽出したが、それ以外の分野でも行うことが挙げられる。これにより、学生の学習向上の一助となることが期待される。

参考文献

- [1] 文部科学省：数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）、文部科学省、入手先<https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm>（参照2023-12-09）。
- [2] 情報処理学会：カリキュラム標準 J17、情報処理学会、入手先<https://www.ipsj.or.jp/annai/committee/education/j07/curriculum_j17.html>（参照2023-07-17）。
- [3] 関谷貴之、松田源立、山口和紀：LDAとIsomapを用いた計算機科学関連カリキュラムの分析、情報処理学会論文誌、Vol.54, No.1, pp.423-434 (2013)。
- [4] 株式会社インセプト：IT用語辞典 e-Words、株式会社インセプト、入手先<<https://e-words.jp/>>（参照2023-08-01）。