

N-004

科目分類支援システムの改善とその応用 An Improvement and an Application of the Course Classification Support System

宮崎 和光[†] 井田 正明[‡] 芳鐘 冬樹[‡] 野澤 孝之[‡] 喜多 一[§]
Kazuteru Miyazaki Masaaki Ida Fuyuki Yoshikane Takayuki Nozawa Hajime Kita

1. はじめに

大学評価・学位授与機構(以下, NIAD-UE と略す)では, 短期大学・高等専門学校卒業生及び専門学校修了者等を対象に, 単位累積加算を基にした学士の学位授与事業を行っている。この制度を利用して学士の学位授与を希望する者は, 各専門分野(専攻区分と呼ばれる)ごとに定められている所定の単位を, 大学の科目等履修生制度などを利用した学修により修得し, それら修得した科目を自らの判断で分類し NIAD-UE に申告しなければならない。それに対し, NIAD-UE では, 申請者による分類の正しさを, 各専門分野の委員が専門委員会において申告された科目のシラバスを読むことで検討している。しかしながら, 近年の申請者数の増大から, この作業には膨大な時間と労力を要しており, 情報技術を活用した科目の分類支援が望まれている。

著者らは, これまで, そのような科目の分類支援を目的として科目分類支援システム(CCS)を試作してきた[宮崎 03]。しかし, これまでに試作されたシステムは, 判定の鍵となるデータベース(DB)において, 各専門用語の出現数が十分には考慮されていないという欠点を持つ。そこで, 本稿では, この点を改善することで, より信頼性の高い情報の提示を目指す。平成 15 年度に現実に申告された科目に適用することで提案手法の有効性を確認する。

2. 科目分類支援システムの改善

2.1 科目分類支援システムの概要

科目分類支援システムとは, 電子化されたシラバスを入力とし, 「新しい学士への途」[¶]に即した科目分類候補を出力するシステムである。ここで科目分類とは, 例えば, 本稿が対象とする専攻区分である「情報工学」区分の場合, 申請者が大学等で修得し NIAD-UE に申告した各科目を表 1 の 1~7 のいずれかの科目区分に分類することをいう。

科目分類支援システムの要は, 過去の科目分類判定を蓄積した科目 DB にある。本来, 科目 DB は, 専門委員会において作成されるべきものであるが, 現実の審査業務に活用されている DB は, 当該シラバスと分類判定結果が「1 対多」に対応していないなど, 十分に整備されているとは言えない。そこで, [宮崎 03] では, 平成 14 年度に, インターネット上に公開されている比較的充実した内容を持つ 13 大学の情報系学科の電子化されたシラバスデータを収集し, 当該シラバスと分類判定結果が「1 対多」に対応する科目 DB を作成した。その後,

表 1: 「情報工学」区分の科目区分。

科目区分 1	情報工学基礎理論に関する科目 [4 単位以上]
科目区分 2	計算機システムに関する科目 [6 単位以上]
科目区分 3	情報処理に関する科目 [6 単位以上]
科目区分 4	情報に関連する科目
科目区分 5	情報工学に関する演習・実験・実習科目
科目区分 6	関連科目 [4 単位以上]
科目区分 7	その他科目

注, 「情報工学」区分の専門科目は A 群(科目区分 1~4)と B 群(科目区分 5)に分かれており, A 群から 30 単位以上, A 群と B 群の合計が 40 単位以上となるように修得しなければならない。さらに, 本稿では, 科目区分 1~5 に分類困難な専門科目を科目区分 0 として分類する。

この科目 DB を加工することで, 各科目区分ごとに, その区分のみに内包されている特徴的な語句の集合である myDB を作成した。具体的には, myDB は, 科目区分 i ($i=0, 1, 2, 3, 4, 6, 7$) に分類されているシラバス内に存在する専門用語集合^{||}から, 科目区分 i に関係しない分類項目に分類されている専門用語集合を取り除くことで作成される。

[宮崎 03]において試作された CCS は, myDB とシラバスのマッチングをとるシステムである。CCS に基づく分類支援は, 次のように実施される。(0) 分類すべき科目のシラバスを [井田 03] が提案する XML 形式のフォーマットに従い電子化し CCS へ入力する。(1) 入力されたシラバスから [湯本 01] の手法により専門用語を抽出する。(2)(1) で抽出された各専門用語と myDB とのマッチングをとり, マッチした用語とその用語に対応する科目区分番号を出力する。

2.2 科目分類支援システムの改善

2.2.1 myDB の改善

myDB は, 各科目区分ごとに, その科目区分のみに含まれている専門用語を集めたものである。しかし, ここでは, その用語の出現数は考慮されていない。本来, ある特定の専門用語が, ある科目区分にとって特徴的な語であるか否かは, その語が, どれだけ多く, その区分に関係するシラバス内に出現したかに依存し変化するはずである。そこで, 本稿では, そのような各専門用語の出現数を考慮した形で myDB を構成し直し, 科目分類支援システムに組み込むことを考える。

そのために, 本稿では, まず初めに, 先の 13 大学のすべてのシラバスから [湯本 01] の手法により専門用語を抽出し, その抽出された各専門用語が, 各科目区分 i ($i=1, 2, 3, 4$) に属する確率を区間推定した。具体的には, ある専門用語が科目区分 i ($i=1, 2, 3, 4$) に属する確率は, 「その用語が含まれているシラバス数」に対する「その用語が含まれており, かつ, 科目区分が i であるシラバ

[†]大学評価・学位授与機構 学位審査研究部

[‡]大学評価・学位授与機構 評価研究部

[§]京都大学 学術情報メディアセンター

[¶]NIAD-UE が行う学士の学位授与制度・申請方法などを詳しく説明した冊子である。

^{||}専門用語は [湯本 01] の手法により抽出される。

ス数」の割合として計算される。各専門用語とこの各科目区分 $i(i=1, 2, 3, 4)$ に対する区間推定値とを組にした DB を以下では $myDB_c$ と呼ぶ。 $myDB_c$ は、これまでの $myDB$ とは異なり、各専門用語の出現数を考慮した形の DB となる。

2.2.2 信頼度を考慮した科目分類支援システム

CCS の $myDB$ を $myDB_c$ に置き換えることを考える。その結果、システムの出力は、「マッチした用語とその用語が各科目区分 $i(i=1, 2, 3, 4)$ に属する確率の区間推定値」となる。以下では、このシステムを信頼度を考慮した科目分類支援システム (Course Classification Support System with confidence; 以下, CCS_c と略す) と呼ぶ。

ところで、専門委員会では、申請者が申告した科目が、各科目区分 (または群) に課せられている単位数を満足しているか否かが議論される。ここでは、このような審査業務の支援に CCS_c を活用することを考える。

このような場合、 CCS では、注目している科目区分とマッチする専門用語の有無に着目し、支援情報を提示していた。しかし既に述べたように、そこでは各専門用語の出現数は考慮されていない。

それに対し、科目区分 i に関係する科目を探しているときの CCS_c は次のように構成される。(0) 申請者が申告しているすべての科目のシラバスと $myDB_c$ とのマッチングをとる。その結果、各シラバスごとに、 $myDB_c$ とマッチした語と、その語の各科目区分に属する確率の区間推定値が得られる。(1) 閾値 $t (0 \leq t < 1)$ を設定する。(2) 科目区分 i へ属する確率の区間推定値の下界値が、 t より大きい科目を提示し、提示された科目を専門委員に判定して頂く。(3) 科目区分 i に課せられている単位数が満足されれば、その科目区分の判定を終える (必要に応じ、他の科目区分または群の判定に移る)。満足されない場合は、閾値を下げ (2) へ戻る。

これにより、各専門用語の出現数を考慮した分類支援が可能となる。

3. 適用事例

平成 15 年度に実際に申請のあった M 大学理工学部の 48 科目および K 大学工学部の 28 科目を対象に CCS_c の有効性を検証する。

3.1 シラバスの直接入力

現在、NIAD-UE には、年間、約 2,000 件にもものぼる申請があり、その結果、全体では、数万ページに達するシラバスを、毎年、取り扱う必要がある。2.1 節で述べたようなこれらすべてのシラバスの XML 化およびそこからの用語抽出は非現実的な作業を伴う。そこで、本稿では、シラバスを直接 CCS 等に入力することを試行した。具体的には、紙ベースで提供されたシラバスをスキャナに通し OCR 処理したものを、直接、 CCS および CCS_c へ入力した。

表 2: CCS_c および CCS の結果。

	科目区分	CCS_c						CCS
		0.6	0.5	0.4	0.3	0.2	0.1	
M 大 学	1	5/8	6/15					4/13
	2	15/17	19/23	20/38				18/24
	3	4/4	5/6					5/8
	4	-	-	-	0/1	1/13		0/3
K 大 学	1	3/3	5/7	6/9				4/4
	2	3/3	6/6	6/9	7/16			5/6
	3	1/1	1/1	3/3				2/4
	4	2/2	3/3	3/3	4/4	5/7	6/21	5/5

3.2 適用結果

結果を表 2 に示す。 CCS の結果は最右列、 CCS_c の結果は中央に閾値ごとに示した**。ここで、“/” の右側の数字は、「各手法がその科目区分に分類されると示唆した科目数」であり、左側の数字は、「示唆された科目のうち、専門委員会においても同じ科目区分に分類されていた科目数」である。例えば、 CCS が科目区分 1 に分類される可能性があるとして示唆した科目は、M 大学の場合 48 科目中 13 科目、 CCS_c の場合は閾値が 0.6 の時点で 8 科目であった。そして、それら示唆された科目のうち、専門委員会においても同一の科目区分に分類されていた科目数は、それぞれ、4 および 5 であった。

CCS_c の結果の各行の最右端に位置する“/”の左側の数字 (太字) は、対象科目のうち、その科目区分に分類されると専門委員会において実際に判定された総科目数である。このことから、 CCS_c では、閾値を下げることで徐々に審査すべき科目候補を増やし、最終的には、全科目を提示するよりも早く、科目区分 1~4 に関係するすべての科目を発見できることがわかる。それに対し、 CCS で同様の結果を得るためには、M 大学の科目区分 3 以外は、専門委員にすべての科目の判定を依頼する必要がある。このことから、 CCS_c は、 CCS よりもより効果的な科目の分類支援が可能であることが確認された。

4. おわりに

本稿では、これまで提案されていた科目分類支援システムの改善を行った。平成 15 年度に現実に申告された科目に適用することで提案手法の有効性を確認した。今後は、実用化に向けさらなる洗練化をはかるとともに、「情報工学」以外の他の専攻区分への適用を検討する予定である。

参考文献

- [宮崎 03] 宮崎和光, 井田正明, 芳鐘冬樹, 喜多 一: 電子化されたシラバスに基づく科目分類支援システムの開発について, 第 2 回情報科学技術フォーラム (FIT2003), pp. 381-382 (2003).
- [井田 03] 井田正明, 宮崎和光, 芳鐘冬樹, 喜多 一: シラバス XML データベースシステム構築に関する考察, 情報処理学会第 65 回全国大会, pp. 247-248 (2003).
- [湯本 01] 湯本紘彰, 森 辰則, 中川裕志: 出現頻度と接続頻度に基づく専門用語抽出, 情報処理学会第 145 回自然言語処理研究会, pp. 111-118 (2001).

**実際のデータは閾値 0.8 から存在するが、紙面の都合上 0.6 から示した。