原著論文

分類候補数の能動的調整を可能にした 学位授与事業のための科目分類支援システムの 提案と評価[†]

宮崎 和光^{*1}・井田 正明^{*1}・芳鐘 冬樹^{*1}・野澤 孝之^{*1}・喜多 一^{*2}

大学評価・学位授与機構（NIAD-UE）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。この制度を利用し学士の学位授与を希望する者は、各専門分野ごとに定められた所定の単位数以上を修得しなければならない。申請者は、自らの判断で修得した科目をNIAD-UEが定める科目区分に合致するように分類・整理し申告する。それに対し、NIAD-UEでは、申請者による分類の正しさを、各専門分野の専門委員が、各科目区分ごとに設定されている例示科目を手がかりに、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。これに関し、各科目区分ごとに、その科目区分のみに含まれている特徴的な用語をキーワードとして抽出することで、専門委員の分類作業を支援するシステムとして、「科目分類支援システム」（CCS）が提案されている。CCSを用いて、ある科目区分に分類される可能性のある科目を探す場合には、その科目区分のみに含まれるとされた用語を含むか否かの基準で判定が行われる。そのため、そのような用語をシラバス内に含む科目は全て同列に扱われる。しかし、NIAD-UEの学位授与事業では必ずしもすべての科目を分類する必要はなく、ある一定の基準を満たした時点で終了することが可能である。そこで本論文では、シラバス内に存在する用語の各科目区分への結び付きの強さを連続値で評価することで、分類を試みている科目区分に帰属する可能性の高い順に申告された科目を専門委員に提示する「分類候補数の能動的調整を可能にした科目分類支援システム」（ACCS）の構築を行う。これにより、専門委員等が自ら能動的に判定すべき科目の候補を調整することが可能になり、分類作業のさらなる負担軽減が期待できる。平成15および16年度に実際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。

キーワード：シラバス、科目分類、学位授与、推薦システム

1. はじめに

大学評価・学位授与機構（NIAD-UE）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。この制度を利用し学士の学位授与を希望する者は、専攻区分と呼ばれる各専門分野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・整理しNIAD-UEに申告しな

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、各専門分野の専門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。

シラバスは学生に対する授業の選択と履修時のガイドのための情報提供を本来の目的とするが[1]、授業内容を最も詳しく示す資料として、大学評価の支援[2]や大学における教育情報の活用[3]などにも積極的に活用されている。また、シラバスは、Webからの自動収集[4]、ならびに、そこからの情報抽出[5]といった情報技術研究の素材としても、近年、特に注目を集めている。これに対し、我々の主眼は、学位授与事業の支援のためにシラバスをその内容に基づき分類することにある。NIAD-UEでは、申請された科目のシラバスを、短期間に、数種類の科目区分に分類することが要求される。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強く望まれている。

[†] Proposal and Evaluation of the Active Course Classification Support System for Degree Awarding - Control of the Number of Candidate Courses in Classification
Kazuteru MIYAZAKI, Masaaki IDA, Fuyuki YOSHIKANE, Takayuki NOZAWA and Hajime KITA

^{*1} 独立行政法人大学評価・学位授与機構
National Institution for Academic Degrees and University Evaluation

^{*2} 京都大学
Kyoto University

る。

このような NIAD-UE におけるシラバスの内容に基づく科目分類は、学位授与事業の持つ審査業務という性質から、申請者の不利益となるような判定を許容できないという性質を持つ。そのため情報技術による支援を行う際には、科目分類の主体はあくまで専門委員会にあり、システムには従役の役割、すなわち推薦システム[6]としての振る舞いが期待されている点に特に注意を払う必要がある。

推薦システムは、一般に、内容に基づくフィルタリング(コンテンツ分析)と協調フィルタリングを利用したシステムに大別される[7,8]。前者は、ユーザの嗜好を記録したユーザプロファイルに基づいて、入力された情報の中から、そのユーザにとって関心があると思われる情報を提示するものであり、後者は、多種多様なユーザがいる場合に、類似した嗜好を持つユーザには、同種の情報を提供するための技術である。NIAD-UE における科目分類では、科目区分がユーザに相当すると考えられる。科目区分は高々数種類であり、また、各科目区分間の類似性を考慮に入れる必要はほとんどないことから、本論文が目指す科目の分類支援では、内容に基づくフィルタリングとの関連性が強いと考える。

内容に基づくフィルタリングを利用した推薦システムは数多く存在する(例えば文献[7,9]など)が、一般に、その成否は対象領域に適合したユーザプロファイルをいかに構成するかにかかっている。このことは、本論文の場合、各科目区分ごとに、その科目区分を特定できる情報をいかに構成するかという問題に帰着される。

NIAD-UE では、そのような科目区分の特定を支援するための情報として、各専攻区分にとって特徴的かつ重要な「専門科目」に対しては、各科目区分ごとに「専門科目の例」を公開している。申請者、専門委員ともに、この例示された科目名を手がかりに未分類の科目の分類を試みる。システムの立場に立てば、この例示科目は、内容に基づくフィルタリングを行う際のユーザプロファイルとなり得るが、例示科目のみでは期待する性能が得られないことが既に報告されている[10]。それに対し、文献[10]では、各科目区分ごとに、その科目区分のみに含まれている特徴的な用語をキーワードとして抽出し、その抽出されたキーワードと申請された科目のシラバスとのマッチングをとるシステムとして、「科目分類支援システム」(CCS)を提案している。

CCS を用いて、ある科目区分に分類される可能性のある科目を探す場合には、その科目区分のみに含ま

れるとされた用語を含むか否かの基準で判定が行われる。そのため、そのような用語をシラバス内に含む科目は全て同列に扱われる。しかし、NIAD-UE の学位授与事業では必ずしもすべての科目を分類する必要はなく、ある一定の基準を満たした時点で終了することが可能である。その意味から、分類を試みている科目区分に帰属する可能性の高い順に、申告された科目をその根拠となる用語とともに専門委員に提示することができれば、専門委員の分類作業の軽減が大いに期待できる。

そこで本論文では、既存の科目分類結果をもとに各用語と各科目区分との結び付きを関係値と呼ばれる統計的な連続値で評価し、[用語, 科目区分 1 に対する関係値, ..., 科目区分 i に対する関係値] (ここで i は科目区分の種類)の組を 1 エントリとする新たなデータベースを構築する。このデータベースと各科目のシラバスとのマッチングをとることで、各科目区分への結び付きの強さを考慮に入れた科目分類支援システムの構築を行う。これにより、専門委員等が自ら能動的に判定すべき科目の候補を調整することが可能になり、分類作業の軽減が期待できる。

ここで、CCS ならびに本論文で提案する手法の出力に分類の根拠となる用語が含まれている点に注意されたい。一般に、シラバスをテキストデータとして捉えれば、文献[11]に述べられているサポートベクターマシン等を利用した手法を用いて、最も適合する科目区分に自動分類し、その結果を専門委員に提示することも可能である。また、文献[12]の確率モデルに基づく技術を用いれば、自動分類した結果に確信度を付与することも可能である。しかしながら、我々の主眼はそのような形での自動分類にはない。NIAD-UE で求められているのは、専門委員の支援である。その意味で、専門委員にも容易に理解できる形、すなわち、分類の根拠となった用語を含めた形で情報を提示できる点が特に重要な意味を持つ。そして、さらに本論文では、その根拠となった用語、各々に対し、各科目区分への結び付きの強さを提示できることで、支援システムとして、さらに優れたものになることが期待される。そのような提案手法の有効性を平成 15 および 16 年度に実際に申告された科目に適用することで確認する。

2. NIAD-UE における科目分類の現状

2.1 単位修得に着目した申請方法の概略

NIAD-UE における学位授与事業の制度・申請方法などの詳細は、NIAD-UE が毎年度発行している「新しい学士への途」[13]に詳しく示されている。ここで

は、その中から、科目分類に直接関係する部分である単位修得に着目した申請方法の概略を述べる。

単位修得の観点で、申請者が第一に行うべきことは、専攻区分の選択である。平成16年度現在、NIAD-UEでは、国語国文学から体育学まで全部で55個の専攻区分を用意している。申請者はこれら55個の中から自らの学修に最も近いと思われる一区分を選択し、その選択した専攻区分に課せられている単位の修得基準である専攻基準を満たす学修を行う必要がある。

表1および表2に示すように、専攻基準は、「専門科目」、「関連科目」および「専攻外科目」の区分欄と「専門科目の例」から構成される。さらに、「専門科目」、「関連科目」の区分欄は、「科目区分」と呼ばれるいくつかの「～に関する科目」に分類され、一部の「科目区分」には最低限必要な単位数が示されている。また、各専攻区分にとって特徴的かつ重要な「専門科目」の科目区分に対しては、分類作業を手助けするための情報として、数個の「専門科目の例」が示されている。平成16年度時点での、「情報工学」区分の「専

門科目の例」は、表2の通りである。

申請者は、修得した科目を、その内容により、各専攻区分に係る「専門科目」、「関連科目」およびそれら以外の「専攻外科目」に自ら分類・整理しなければならない。これは、上述の「情報工学」区分の場合は、表1の1～7のいずれかの科目区分に修得した科目を申請者自らが分類することを意味する。このように申請が自己申告を基本としているため、場合によっては、不適切な科目分類が行われる可能性があり、NIAD-UEにおいて、その適切性の確認が行われている。

2.2 現状の科目審査プロセスの概要

専門科目の判定は、各分野に精通した専門家による判定が必要である。NIAD-UEでは、そのような専門性の高い科目分類の確認を支援するために、平成12年度から、過去の科目分類の判定結果をデータベース化し、専攻基準の判定に活用している。このデータベースは科目データベースと呼ばれており、現状では、ひとつのシラバスは、どれかひとつの科目区分にのみ分類されるように作られている。

表1 「情報工学」区分の科目区分

専門科目 (40単位以上)	A群 (30単位以上)	情報工学基礎理論に関する科目(4単位以上) 計算機システムに関する科目(6単位以上) 情報処理に関する科目(6単位以上) 情報に関連する科目	科目区分1 科目区分2 科目区分3 科目区分4
	B群	情報工学に関する演習・実験・実習科目	科目区分5
関連科目 (4単位以上)	工学の基礎となる科目 工学及び周辺技術等に関する科目		科目区分6
専攻外科目	「専門科目」及び「関連科目」以外の科目		科目区分7

[平成16年度版「新しい学士への途」より抜粋]

表2 「情報工学」区分の「専門科目の例」

○情報工学基礎理論に関する科目(科目区分1)

データ構造、言語理論とオートマトン、アルゴリズム、計算理論、符号理論、スイッチング回路理論、信号処理理論、論理学、情報工学、情報数学、離散数学(グラフ理論、組合せ論)、計画数理(OR、待行列)など

○計算機システムに関する科目(科目区分2)

プログラミング、プログラム言語、コンパイラ、オペレーティングシステム、デジタル回路、計算機構成、計算機アーキテクチャ、ネットワーク、情報セキュリティ、分散処理、ソフトウェア工学、データベースシステム、性能評価など

○情報処理に関する科目(科目区分3)

数値解析、人工知能、知識工学、エキスパートシステム、自然言語処理、音声処理、画像処理、図形処理、パターン認識、シミュレーション、通信処理(プロトコル)、グループウェア、マルチメディア、ヒューマンインターフェース、コンピュータグラフィックス、ロボティクスなど

○情報に関する科目(科目区分4)

集積回路、電子回路、システム制御理論、通信網、最適化論、計測工学、生体情報処理など

[平成16年度版「新しい学士への途」より抜粋]

図1 現状の科目審査プロセスの例
(文献[10]より引用)

図1に、科目データベースを用いた専攻基準判定のための科目審査プロセスの一例を示す。まず申請者による分類と科目データベースとのマッチングをとり、一致するものは「可」とする。図1の例では、「アルゴリズムとデータ構造」「コンパイラ」がこれに相当する。一方、不一致となった場合には、機械的に申請者による科目分類を、科目データベースの分類に修正（移動）する。図1の例では、「計画数学」がこれに相当する。

それに対し、過去に判定例が存在しない科目（初出科目）については、専門委員会において、申請者による科目分類の正しさを、その科目のシラバスを読むことで検討している。図1の例では、「数理経済学」がこれに相当する。初出科目に対する判定作業は、基本的には対応するシラバスをすべて読むことで行われており、近年の申請者数の増大から、その作業には膨大な労力と時間を要している。特に、専門委員会では、表2に示すような「専門科目の例」が存在する科目区分に対する分類作業に多くの時間が費やされる。そこで、本論文では、「専門科目の例」が存在する科目区分を対象に、専門委員の作業量軽減を目指す。

2.3 科目分類支援システム

現在、このような科目分類に係る判定作業を支援するシステムとして、「科目分類支援システム」(Course Classification Support system; 以下、CCSと略す)[10]が存在する。CCS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、ひとつのシラバスが、複数の科目区分に分類されることが許容されていないなど、システム化に際して、不都合な点が存在する。そこで、文献[10]では、「情報工学」区分を対象として、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（全1,010科目）

を収集し、それらすべてのシラバスを著者らが熟読することで、ひとつのシラバスが、複数の科目区分に分類されることを許容した科目DBを作成し、既存の科目DBの不十分さを補っている。

その後、文献[10]では、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用語の集合である $myDB = \{myDB(1), \dots, myDB(i)\}$ (i は科目区分の種類) を作成している。各 $myDB(i)$ は、科目区分 i に分類されているシラバス内に存在する用語集合¹から科目区分 i と無関係な区分に分類されている用語集合を取り除くことで作成される。ここで先の13大学の場合、「情報工学」区分の「専門科目」A群の各科目区分 i ($i = 1, 2, 3, 4$) に対応する $myDB(i)$ に登録されている用語の数は、それぞれ954, 1318, 1127, 754である。

CCSは、 $myDB$ とシラバスのマッチングをとるシステムである。科目区分 i に帰属する科目を探しているときのCCSは、以下のように構成される。

step 0: 申請者が申告している科目のシラバスと $myDB(i)$ とのマッチングをとる。その結果、各シラバスごとに、 $myDB(i)$ とマッチングした用語の集合が得られる。

step 1: step 0で得られた集合が空でない科目を無作為にひとつ選択する。

step 2: 選択された科目のシラバスに、step 0で得られた用語集合を付して専門委員に提示し、判定作業を行って頂く。

step 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、step 0で得られた集合が空でない科目の中で未判定な科目を無作為にひとつ選択し、step 2へ戻る。

CCSに基づく分類支援の特長は、 $myDB(i)$ とのマッチングの有無により、専門委員に提示する科目集合を限定している点にある。この限定により、従来の全科目を判定する場合に比べ、専門委員の作業量の軽減が期待できる。しかし、無作為に科目を選択するプロセスに対しては、さらなる効率化が期待されること、ならびに判定すべき科目集合の大きさを能動的に専門委員等が調整することができないこと、が欠点として挙げられる。

3. 科目分類支援システムの改善

3.1 $myDB$ の改善

前節で述べたCCSが持つ欠点を克服するために、

¹ 各用語は文献[14]の手法により抽出される。

本論文では、各用語と各科目区分との結び付きを関係値と呼ばれる連続値で評価することを考える。この関係値を利用すれば、判定すべき科目の集合を段階的に増やすことが可能となる。その結果、CCSの欠点も克服される。

ところで、CCSの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用語を集めたものである。しかしそこでは、その用語と、その用語が含まれている科目数との関係は考慮されていない。本来、ある特定の用語が、ある科目区分にとって特徴的な語であるか否かは、その用語が、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。そこで本論文では、このような各用語と、その用語が含まれている科目数との関係を考慮した形でmyDBを構成し直すことを考える。

そこで、まずはじめに、先の13大学のすべてのシラバスから文献[14]の「接続頻度に基づく手法」により用語を抽出し、その抽出された各用語が、各科目区分*i*に関係するシラバス内に含まれる確率を区間推定した。ある用語が科目区分*i*と関係するシラバス内に含まれる確率の信頼区間は、「その用語が含まれているシラバス数」(n)に対する「その用語が含まれており、かつ、科目区分が*i*であるシラバス数」(x)の割合として以下の式から求められる[15]。

$$\frac{\frac{x}{n} + \frac{Z_{\gamma}^2}{2n} \pm \frac{Z_{\gamma}}{\sqrt{n}} \sqrt{\frac{x}{n} \left(1 - \frac{x}{n}\right) + \frac{Z_{\gamma}^2}{4n}}}{1 + \frac{Z_{\gamma}^2}{n}} \quad (1)$$

ここで Z_{γ} は、信頼度95%のとき1.96、信頼度99%のとき2.58となる正規分布の両側パーセント点である。なお、4節では、 $Z_{\gamma}=1.96$ (信頼度95%)とした。

この計算結果、すなわち、「ある用語が、科目区分*i*に関係するシラバス内に存在する確率の区間推定値」を各科目区分ごとに関係値としてとりまとめ、[用語、科目区分1に対する関係値、..., 科目区分*i*に対する関係値] (i は科目区分の種類)の組を1エン트리とするDBを構築する。以下では、このDBをmyDB_cと呼ぶ。myDB_cは、これまでのmyDBとは異なり、各用語と、その用語が含まれている科目数との関係を考慮に入れたDBとなる。また関係値を参照することで、各用語の各科目への結び付きの強さを連続的に評価することもできる。

「アルゴリズム」という用語を取り上げ具体的な関係値の例を示す。「アルゴリズム」は、先の13大学の1,010科目のシラバスのうち95科目のシラバスに登場し、その中で、科目区分1, 2, 3, 4に分類されていたシラバスは全部で47, 26, 21, 1科目であった。従っ

て、信頼度を95%に設定した場合には、科目区分1に対しては、(1)式より信頼区間の上側信頼限界および下側信頼限界は、それぞれ0.594, 0.396となる。これを(0.396, 0.594)と書けば、科目区分1, 2, 3, 4に対する関係値は、各々、(0.396, 0.594), (0.194, 0.371), (0.149, 0.314), (0.002, 0.057)となる。

3.2 分類候補数の能動的調整を可能に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の各科目区分に対する関係値」となる。以下では、このシステムを「分類候補数の能動的調整を可能にした科目分類支援システム」(Active Course Classification Support system; 以下、ACCSと略す)と呼ぶ。

科目区分*i*に帰属する科目を探しているときのACCSは、以下のように構成される。

step 0: 申請者が申告している科目のシラバスと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とマッチングした用語と、その語の各科目区分に対する関係値が得られる。

step 1: 閾値 t ($0 \leq t < 1$)の初期値を設定する。

step 2: step 0で得られた関係値を参照し、科目区分*i*に関係するシラバス内に存在する確率の信頼区間の下側信頼限界が t 以上となっている用語を含む科目のシラバスを、その用語とともに専門委員に提示し、判定作業を行って頂く。

step 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step 2へ戻る。

ところで、(1)式を用いた区間推定では、サンプル数 n 、すなわち、「その用語が含まれているシラバス数」が大きくなればなる程、推定精度が向上し、区間幅が狭まる。したがって、推定の中央値が同じ値のもの同士を比較した場合、下側信頼限界が小さいもの程、より信頼がおけない推定結果を提示していることを意味する。上のstep 2において、下側信頼限界を用いているのは、このようなサンプル数の違いによる差異を判定結果に適切に反映させるためであるが、将来的には、下側信頼限界と上側信頼限界の併用も検討する予定である。

またシステムの出力に分類の根拠となる用語が含まれている点には特に注目されたい。一般に、シラバスをテキストデータとして捉えれば、文献[11]に述べられている手法等を用いて、最も適合する科目区分に自

動分類することも可能である。しかしながら、はじめに述べたように、我々の主眼はそのような自動分類にはない。NIAD-UEで求められているのは、専門委員の支援である。その意味で、専門委員にも容易に理解できる形、すなわち、分類の根拠となった用語を含めた形で情報を提示できる点が特に重要な意味を持つ。

3.3 閾値と用語数との関係

これまでのmyDBを用いた分類支援では、マッチングの対象となる用語数が固定されていた。それに対し、myDB_cを用いたACCSでは、閾値を変化させることで、動的にマッチングの対象となる用語数を変化させられる。

この変化の様子を例示するために、先の13大学の1,010科目のシラバスから抽出された用語（全部で16,452語）を対象に、閾値を0.7から0.1まで0.1刻みで変化させたときの用語数の変化を調べた。具体的には、各科目区分ごとに、その科目区分に関係するシラバス内に存在する確率の信頼区間の下側信頼限界が閾値以上である用語の数を数え上げ、各閾値ごとに集計した。結果を表3中央に示す。また参考までに、myDBに登録されている用語数を表3最右列に示した。このことから、myDB_cを用いることで、マッチングの対象となる用語の数を動的に変化させられることがみてとれる。

これにより、当初の目的である、各科目区分ごとに、その科目区分に分類される可能性が高い科目をその根拠となる用語とともに順次提示することが可能になる。さらに、閾値を調整することで、システムから提示される科目数を専門委員等が、直接、コントロールすることも可能となる。

4. 適用事例

平成15および16年度に「情報工学」区分に実際に申請のあった4大学（M、K、O、H）の各科目（48、28、56および30科目）を対象にACCSの有効性を検証する。

4.1 シラバスの直接入力

ところで、現在、NIAD-UEには、年間、約2,000

表3 閾値と用語数との関係

科目 区分	myDB _c							myDB
	0.7	0.6	0.5	0.4	0.3	0.2	0.1	
1	10	31	100	198	414	1605	1927	954
2	11	51	160	330	704	2095	2374	1318
3	4	23	67	152	374	1537	1796	1127
4	2	17	44	75	186	967	1113	754

件に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。これらすべてのシラバスが文献[16,17]で提唱されているシラバスデータベース（シラバスDB）に登録されていることが理想であるが、そのための作業であるシラバスのXML化は非現実的な作業量を要求する。そこで、本論文では、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ACCSへ入力した。なお、OCRソフトは定評のある市販のものを利用した。

これによりシラバスのXML化にかかる労力が大幅に削減される。4.2および4.3節では、OCRソフトの認識結果をそのまま受け入れた場合の結果を示す。その一方で、このことは、入力信号に、誤認識された文字や改行により不適切に分断された文字等が混入する可能性を含む。そのようなOCRソフトの誤認識については4.4節で議論する。

4.2 適用結果

紙ベースで提供される先の4大学のシラバスをOCR処理したものをCCSおよびACCSに入力した。すなわち、全部で162科目（＝48＋28＋56＋30科目）のシラバスを表3中央に示した用語数を持つmyDB_cとマッチングさせた。結果を表4に示す。CCSの結果は最右列、ACCS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唆した科目は、M大学の場合48科目中13科目、ACCSの場合は、例えば閾値が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科目であった。また、各大学名の下に数字は、対象とした科目数である。

ACCS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、ACCS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、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

表 4 ACCS および CCS の結果

	科目 区分	ACCS							CCS
		0.7	0.6	0.5	0.4	0.3	0.2	0.1	
M 大 学 48	1	3/3	5/8	6/11					4/13
	2	5/6	14/16	20/27					18/24
	3		4/4	5/6					5/8
	4					0/4	1/7		0/3
K 大 学 28	1	2/2	3/3	5/7	6/9				4/4
	2	2/2	3/3	6/6	6/10	7/18			5/6
	3		1/1	1/1	3/3				2/4
	4	1/1	2/2	3/3	4/4	5/5	5/7	6/20	5/5
O 大 学 56	1	3/4	6/7	11/14	15/20	16/27	17/50	18/56	8/12
	2	5/6	10/12	14/20	14/26	16/46			13/15
	3	0/1	1/3	2/6	2/13	3/29			1/4
	4					0/3	3/22	4/44	0/4
H 大 学 30	1	2/4	4/7	4/9	5/11				4/5
	2	2/2	3/3	4/8					3/9
	3		1/4	3/11	4/14				3/8
	4	1/1	3/4	4/5	7/8	7/9	12/20		5/7

門委員にすべての科目の判定を依頼する必要がある。

ACCSにおいても、O大学の科目区分1においては、専門委員にすべての科目の判定を依頼する必要がある。しかしこれは例外的である。というのも、O大学の科目区分1において最後に（18番目に）発見された科目の名称は、「パターン工学特論」であり、その内容は、「パターン情報のセンシング」といった、本来、科目区分3であると判断されるべき内容を持つものである。この科目に対し、申請者は、科目区分1に分類していた。専門委員会においては、（主として時間的制約から）最終的な合否にクリティカルに影響しない部分では、シラバスを熟読せずに、申請者の分類で可とする場合がある。この科目も、このルールにより申請者の分類を修正しなかったものである。このような分類誤りの未修正は、正確な科目データベースの構築の大きな妨げとなる。今後は、専門委員会において、申請者の分類のみならず、本システムの分類結果を並記することで、このような誤りを未然に防ぐ効果も期待できると考える。

なお、M大学よりもO大学は、より小さい値の閾値を必要としている。これはO大学のシラバスは内容が乏しい（全体で数10語程度）ものが多く、その結果、用語のマッチングが生じにくいことが主たる原因である。また、科目区分4は全体的に閾値の小さい部分でのマッチングが多い。これは、表3からもわかる通り、科目区分4に関係する用語の数が他の区分に比べ

少ないためである。

4.3 専門委員の負荷軽減量の比較

本節で扱っている4大学のシラバスは、各大学ごとに異なる4名の申請者から申請されたものである。したがって、各大学ごとに、表1、A群の各科目区分に課せられている単位数の条件、すなわち、「科目区分1は4単位以上、科目区分2および3は6単位以上」を満足する必要がある。ここでは、これら4名の申請者に対し、科目区分1、2、3の順番で専門委員が単位数の確認を行った場合を想定し、CCSとACCSの比較を行う。ちなみに、本節で議論の対象とされる科目はすべて2単位である。そのため、各科目区分に課せられている条件は、「科目区分1は2科目以上、科目区分2および3は3科目以上」と読み換えることが可能である。

表5に、科目区分1、2および3の条件を満足するまでにシステムから提示された科目の総数を示す²。この値が小さければ小さいほど、より専門委員の負荷軽減量が大きい手法であると言える。

CCSでは、K大学およびO大学では、すべての科目の判定が必要とされる。これは、K大学およびO大学では、科目区分3に帰属する科目を3科目以上専門委

² CCSには、無作為に提示する科目を選択するプロセスが含まれているので、システムが最悪の振る舞いをした場合を想定し評価している。

表5 科目区分1, 2, 3の基準を満足するために必要とされた科目数

大学名 (総科目数)	ACCS	CCS
M 大学 (48)	13	45
K 大学 (28)	8	28
O 大学 (56)	39	56
H 大学 (30)	18	22
合計 (162)	78	151

員に提示することができないため、最悪の場合には、すべての科目の判定が要求されることに起因する。M大学およびH大学では、そのようなことはなく、科目区分1, 2, 3の順番に提示された科目を順次判定すれば基準が満たされる。

これに対し、ACCSでは、大幅に少ない科目数で判定を終えている。例えば、M大学では、科目区分1, 2, 3は、それぞれ閾値0.7, 0.7, 0.6, の時点で提示されていた科目数、すなわち、合計13(=3+6+4)科目の判定に必要な単位数を満足している。単純にすべてを合計すれば、全162科目のうち、CCSでは151科目の判定が必要であるのに対し、ACCSでは78科目の判定で済む。これは、従来に比べ、約半分の時間で科目の判定を終えられることを意味する。

4.4 シラバスDBとの連携結果

OCRソフトによるシラバスの電子化は非常に簡便

であるが、文字認識の信頼性の意味では、少々不安が残る。また、そのまま用いると、科目の内容と直接関係がない項目まで含んでしまうという問題もある。そこで文字認識の誤りを手作業で修正したものを、文献[16,17]で提唱されているシラバスDBに登録し、シラバスDBからのデータに基づき判定した場合との比較を次に行う。

シラバスDBには、さまざまな項目が登録されているが、ここでは、各科目の内容と密接に関連する記述が多いと思われる「科目名」「授業概要」「授業計画」「履修により達成される目標」フィールドのみを取り出し、CCSおよびACCSへの入力とした。結果を表6に示す。

ほぼ期待通りの結果が得られた。すなわち、OCRの認識不足による判定誤りが、手作業による文字認識の誤りを修正したシラバスDBを使うことで皆無となった。例えば、本来ひとかたまりであるべき語句が、文字間の空白が大きい、あるいは、用語の途中で改行されている等の理由から、ひとつのまとまった用語として出力されないケースや、純粋に本来とは異なる文字に誤認識してしまうケースが皆無となった。これによる改善、すなわち、シラバスを直接入力したときに比べ、より高い閾値でその科目を正しく判定できるケースは、K, Oの2大学でみられ、それぞれ、3科目ずつ存在した。

逆に、M大学の科目区分2およびH大学の科目区分

表6 シラバスDBとの連携結果

	科目 区分	ACCS						
		0.7	0.6	0.5	0.4	0.3	0.2	0.1
M 大 学 48	1	3/3	5/7	6/10				
	2	8/18	15/15	18/23	20/31			
	3		4/4	5/7				
	4					0/1	1/19	
K 大 学 28	1	3/3	3/3	6/8				
	2	2/2	3/3	6/6	6/10	7/18		
	3		2/2	2/2	3/3			
	4	1/1	2/2	3/3	4/4	5/5	6/8	
O 大 学 56	1	3/3	6/7	12/15	15/20	16/30	17/50	18/56
	2	5/6	10/12	14/20	14/27	16/45		
	3	0/1	1/3	3/7				
	4		1/1	1/1	1/1	1/4	4/21	
H 大 学 30	1	2/4	4/7	4/8	5/9			
	2	2/3	2/4	4/12				
	3		1/5	3/9	3/12	4/20		
	4	1/1	2/2	3/3	5/5	6/6	11/21	12/29

3 および 4 では、性能の悪化、すなわち、シラバスを直接入力したときに比べ、より低い閾値にならないとその科目を正しく判定できないケースが見られた。これは、シラバス DB からの入力を「科目名」「授業概要」「授業計画」「履修により達成される目標」に限定した影響である。M 大学では、先の 4 フィールド以外の「その他」フィールド、H 大学では、「他の科目との関連性」フィールドにその科目を特徴付ける重要な語が含まれていた。これを除外してしまったために、このような性能の悪化がみられた。この悪化は、M 大学では 2 科目、H 大学では 4 科目存在した。シラバスは大学ごとに特徴ある書き方をしていることが多く、シラバス DB を活用する際には、このような重要なフィールドを科目分類支援システムへの入力時に欠落させないことが重要である。

最後に、シラバス DB を用いた場合の方が、提示される科目数が多くなる場合がある。これは、文字認識の誤りを修正したことで、myDB_c とマッチングする用語が単純に増加したためである。

5. おわりに

本論文では、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 および 16 年度に実際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。特にこれまで未考慮であったシラバスと科目区分の結び付きの度合を考慮した形での科目分類候補の提示が可能になったことで、専門委員の作業量の軽減により貢献することがわかった。

今後は、実用化に向けさらなるシステムとしての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が重要であると考え、この場合、まず第一に、その専攻区分に合致した myDB_c を作成しなければならない。myDB_c の作成はある程度の専門性を要するが、一から構築するのは困難な場合も考えられる。そこで、初期には、例示科目のみを myDB_c に登録しシステム構築し、判定結果が増えるにつれ、それを、随時、myDB_c に反映させてゆくのが現実的な方策であると考え、

NIAD-UE で行われているような専門性の高い大量の科目をシラバスのみを考慮して短期間で集中的に分類することは、他に類をみない特徴的な業務であるが、単に科目を分類するという意味では、編入学で新たな学生を受け入れる場合に普通に要求される。また、米国のリージェント大学[18]や韓国の KEDI (Korean Educational Development Institution) に設置されている学点銀行[19]などでは、同様の分類が日常の業務として行われている。今後は、これらの機関

等との連携なども視野に入れつつ、本システムの普及に努めたいと考えている。

謝辞

本研究を遂行するにあたりデータ整理に多大なご協力を頂きました吉田恭子氏に謝意を表します。

参考文献

- [1] 井下理：シラバスの意味と機能、大学セミナーハウス (編)、大学力を創る：FDハンドブック、東信堂、pp.62-72 (1999)。
- [2] 野澤孝之、井田正明、芳鐘冬樹、宮崎和光、喜多一：シラバス文書のクラスタリングに基づくカリキュラム分析システムの構築、情報処理学会論文誌、Vol.46, No.1, pp.289-300 (2005)。
- [3] 山田信太郎、伊東栄典、廣川佐千男：自動収集した Web シラバスデータの分析と考察、情報科学技術フォーラム 2002 (FIT2002)、一般講演論文集第 4 分冊 (N-32)、pp.301-302 (2002)。
- [4] 山田信太郎、松永吉広、伊東栄典、廣川佐千男：Web シラバス情報収集エージェントの試作、電子情報通信学会和文論文誌 D-I, Vol.J86, No.8, pp.566-574 (2003)。
- [5] 伊東栄典、松永吉広、山田信太郎、廣川佐千男：Web シラバスからの DB 構成、2003 年度人工知能学会全国大会 (JSAI2003) 1D4-08 (2003)。
- [6] P. Resnick and H. Varian. Recommender Systems, introduction to special section of Communications of the ACM, March 1997, vol.40(3) (1997)。
- [7] 寺野隆雄：Web 上の情報推薦システム、情報処理、Vol.44, No. 7, pp.696-701 (2003)。
- [8] 李鵬、山田誠二：帰納学習を用いた映画推薦システム、2004 年度人工知能学会全国大会、3H2-03 (2004)。
- [9] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. Combining content-based and collaborative filters in an online newspaper. In Proceedings of the ACM SIGIR '99 Workshop on Recommender Systems: Implementation and Evaluation, Berkeley, California, August 1999.
- [10] 宮崎和光、井田正明、芳鐘冬樹、野澤孝之、喜多一：電子化されたシラバスに基づく学位授与事業のための科目分類支援システムの試作、情報処理学会論文誌、Vol.46, No.3, pp.782-791 (2005)。
- [11] 高須淳宏、相原健郎：テキスト分類における訓練データと性能の実験的考察、NII Journal, No.6, pp.1-8 (2003)。
- [12] 佐藤健吾：確率モデルによる Web ページ推奨エンジン、IPA 平成 13 年度成果報告集、第 2 版、情報処理振興事業協会、(2002)。
- [13] 平成 16 年度版新しい学士への途、http://www.niad.ac.jp/sub_gakuo/siryo/new/new_gakushiH16.pdf
- [14] 中川裕志、森辰則、湯本紘彰：出現頻度と接続頻度に基づく専門用語抽出、自然言語処理、Vol.10, No.1, pp.27-45 (2003)。
- [15] 泉信一、近藤基吉、穂刈四三二、永倉俊充編集、共立数学公式附函数表改訂増補、共立出版株式会社、

- pp.439 (1969).
- [16] 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).
- [17] 井田正明, 野澤孝之, 芳鐘冬樹, 宮崎和光, 喜多一: シラバスデータベースシステムの構築と専門教育課程の比較分析への応用, 大学評価・学位研究, No.2, pp.85-97 (2005).
- [18] 森利枝: リージェント大学における評価のシステム- 学習とクレジットの評価を中心に -, 学位研究, No.10, pp.107-129 (1999).

- [19] KEDI (Korean Educational Development Institution), <http://eng.kedi.re.kr>

(2005年3月31日 受付)

(2005年7月25日 採録)

[問い合わせ先]

〒187-8587 小平市学園西町1-29-1

独立行政法人大学評価・学位授与機構 学位審査研究部

宮崎和光

TEL: 042-353-1834

FAX: 042-353-1861

E-mail: teru@niad.ac.jp

著者紹介

みやざき かずてる
宮崎 和光 [非会員]

平成8年東京工業大学大学院総合理工学研究科知能科学専攻博士後期課程修了, 同年同大学大学院総合理工学研究科助手, 平成10年から11年にかけて同大学大学院総合理工学研究科リサーチアソシエイト, 平成11年学位授与機構審査研究部助教授を経て平成12年より大学評価・学位授与機構学位審査研究部助教授, 機械学習, 特に強化学習に関する研究に従事, 博士(工学), 人工知能学会, 計測自動制御学会, 日本機械学会, 日本高等教育学会等の会員。

いわた まさひろ
井田 正明 [正会員]

平成2年京都大学大学院工学研究科精密工学専攻修士課程修了, 同年京都大学工学部助手, 平成14年より大学評価・学位授与機構評価研究部助教授, システム工学, 評価と意思決定, 数理計画法, データベースに関する研究に従事, 博士(工学), 情報処理学会, 計測自動制御学会等の会員。

よしかわ ふゆき
芳鐘 冬樹 [非会員]

平成14年東京大学大学院教育学研究科総合教育科学専攻博士課程単位取得退学, 同年大学評価・学位授与機構評価研究部助手, 研究コミュニティの構造分析, 専門用語の自動抽出手法に関する研究に従事, 修士(教育学), 情報処理学会, 日本図書館情報学会の会員。

のざわ たかゆき
野澤 孝之 [非会員]

平成14年東京工業大学大学院総合理工学研究科知能システム科学専攻博士課程修了, 平成14年から15年にかけて京都産業大学客員研究員, 平成15年より大学評価・学位授与機構評価研究部助手, 非線型力学, 複雑系の理論とシミュレーション, ネットワーク構造分析に関する研究に従事, 博士(理学), 電子情報通信学会会員。

きまた はじめ
喜多 一 [非会員]

昭和57年京都大学工学部卒業, 昭和62年京都大学大学院工学研究科博士後期課程研究指導認定退学, 昭和62年京都大学工学部助手, 平成9年東京工業大学大学院総合理工学研究科助教授, 平成12年大学評価・学位授与機構教授を経て平成15年京都大学学術情報メディアセンター教授, 社会システム, ニューラルネットワーク, 進化的計算, エージェントシミュレーションの研究に従事, 工学博士, 電気学会, 電子情報通信学会, 計測自動制御学会, システム制御情報学会等の会員。

Proposal and Evaluation of the Active Course Classification Support System for Degree Awarding - Control of the Number of Candidate Courses in Classification

by

Kazuteru MIYAZAKI, Masaaki IDA, Fuyuki YOSHIKANE, Takayuki NOZAWA and Hajime KITA

Abstract :

The National Institution for Academic Degrees and University Evaluation is engaged in the awarding of academic degrees based on the accumulation of credits. These credits must be classified according to pre-determined criteria for the chosen disciplinary field. This work is carried out by the sub-committees within the *Committee of Validation and Examination for Degrees*, whose members are well-versed in the syllabus of each course. The number of applicants is increasing yearly so that a course classification system supported by information technology is strongly desired. We know the *Course Classification Support system* (CCS), which is a prototype system to support the course classification for the awarding of degrees in NIAD-UE. CCS recommends several courses to the *Committee of Validation and Examination for Degrees* at the same time. It cannot give any priority how each course belongs to the class. In this paper, we propose the *Active Course Classification Support system* (ACCS) that can give the *Committee of Validation and Examination for Degrees* the priority which course fits more than the others. According to the priority given by ACCS, the committee can actively control the number of courses that are to be judged their suitability for each class. We show its effectiveness in the classification of courses that were actually submitted in the fiscal year 2003 and 2004.

Keywords : syllabus, course classification, degree awarding, recommender system

Contact Address : **Kazuteru MIYAZAKI**

Faculty of Assessment and Research for Degrees, National Institution for Academic Degrees and University Evaluation,

1-29-1, Gakuennishimachi, Kodaira-city, Tokyo, 187-8587, Japan

TEL : 042-353-1834

FAX : 042-353-1861

E-mail : teru@niad.ac.jp