

## 集合知による知識体系構築のための意見集約手法

矢吹 太朗<sup>†</sup> 森田 武史<sup>††</sup> 増永 良文<sup>†††</sup>

<sup>†</sup> 千葉工業大学社会システム科学部  
〒 275-0016 千葉県習志野市津田沼 2-17-1

<sup>††</sup> 青山学院大学社会情報学部  
〒 252-5258 神奈川県相模原市中央区淵野辺 5-10-1

<sup>†††</sup> お茶の水女子大学名誉教授

E-mail: <sup>†</sup>taro.yabuki@it-chiba.ac.jp, <sup>††</sup>t\_morita@si.aoyama.ac.jp, <sup>†††</sup>yoshi.masunaga@gmail.com

**あらまし** 大学の学部や学科の授業のシラバスに掲載された情報をもとにして、その組織の教育内容を表現する知識体系 (Body of Knowledge, BOK) を構築する方法を提案する。これは、集合知を活用して BOK を構築しようとする試みの一種である。集合知の活用には、集約のメカニズムが必要だと考えられているが、本手法における集約メカニズムはクラスタ分析である。この手法は、計算機科学のためのカリキュラムである Computing Curricula 2001 を構成する授業から、計算機科学のための BOK である CS Body of Knowledge を再現する試みで用いられている。本稿では、この手法を用いて社会情報学の BOK 構築を試みる。構築された BOK を、大学教員が評価したり、現在人手によって構築が進められている BOK と比較することによって、提案手法の有効性を議論する。

**キーワード** 知識体系, 集合知, シラバス, クラスタリング, クラスタ分析

## Opinion Aggregation Method for BOK Construction by Collective Intelligence

Taro YABUKI<sup>†</sup>, Takeshi MORITA<sup>††</sup>, and Yoshifumi MASUNAGA<sup>†††</sup>

<sup>†</sup> Faculty of Social Systems Science, Chiba Institute of Technology  
2-17-1, Tsudanuma, Narashino-shi, Chiba 275-0016, Japan

<sup>††</sup> School of Social Informatics, Aoyama Gakuin University  
5-10-1 Fuchinobe, Chuo-ku, Sagami-hara-shi, Kanagawa 252-5258, Japan

<sup>†††</sup> Professor emeritus of Ochanomizu University

E-mail: <sup>†</sup>taro.yabuki@it-chiba.ac.jp, <sup>††</sup>t\_morita@si.aoyama.ac.jp, <sup>†††</sup>yoshi.masunaga@gmail.com

**Abstract** We propose BOK (Body of Knowledge) construction method that uses syllabuses of a department or a faculty of a college. This method is a kind of trials that harness collective intelligence to construct BOK. Cluster analysis is introduced as an aggregation mechanism that is thought to be essential to harness collective intelligence. We have tried this method to reproduce Computer Science Body of Knowledge from courses of Computing Curricula 2001. In this paper, we try to construct Social Informatics Body of Knowledge. The effectiveness of our method is discussed based on examination of the result by professors and comparison with BOK being made by another method.

**Key words** body of knowledge, collective intelligence, syllabus, clustering, and cluster analysis

### 1. はじめに

知識体系 (Body of Knowledge, **BOK**) は特定の分野を構成する概念をまとめたものである。代表的な BOK の 1 つに、計算機科学のカリキュラム, Computing Curricula 2001 Computer Science (**CC2001**) とともに発表された計算機科学の BOK, CS Body of Knowledge (**CSBOK**) がある [1].

#### 1.1 BOK の重要性

BOK には、大学教育における非常に重要な 2 つの役割がある。カリキュラムの作成と教育内容の把握である。

カリキュラムの作成には教育機関の個別の事情 (教員数や教員の専門分野) が影響するが、教育内容があらかじめ BOK によって規定されていれば、個別の事情を考慮しながらも標準的

な教育を行える。たとえば文献[1]では、CSBOKに対応したカリキュラムが複数提案されている。

教育内容の把握は、昨今需要が高まっている「大学教育の質保証」に必須の活動である。教育内容の把握には、個別の事情をそのまま反映しているカリキュラムよりも、その背後にあるBOKの方が適している。

確定したBOKがない場合でも、カリキュラムを基に暫定のBOKを構築できれば、それは有用である。カリキュラム作成時には、暫定のBOK上での考察をカリキュラムにフィードバックできるし、カリキュラムが完成した後は、暫定のBOKで教育内容を把握できるからである。

## 1.2 BOK 構築の試み

情報科学のような比較的伝統のある分野では、文献[1]のようにカリキュラムとBOKの両方が揃っているが、新生学問分野（例：社会情報学）では、カリキュラムはあってもBOKはなく、BOKの裏付けがないままに教育が行われることがある。この原因として、CSBOKを構築したIEEE Computer SocietyとACMの統合委員会のメンバのような、分野全体に精通する有識者がいないことが挙げられる。

そこで、集合知によるBOK構築が試みられてきた。**WikiBOK**である[2]。WikiBOKは、木構造で表現されたBOK（BOK木）を、その場で簡単に編集できるウェブページであるWikiを使って、不特定多数の集団で構築しようという試みである。WikiBOKは有望な試みではあるが、次のような問題が指摘されてきた。

（1）カリキュラムとの対応が考慮されていない。

（2）特定のテーマのBOK中での重要度を判断するのが難しい。BOK木のノードは上から**エリア・ユニット・トピック**と呼ばれるが、たとえば「統計学」をBOKに導入しようとするときに、それをどの階層に入れるべきか、「記述統計」と「推測統計」に分けるべきかといった判断は難しい。

（3）木構造を直接編集できるユーザインタフェースでは集合知が機能しない。たとえば、ある参加者がエリア（BOKの第1階層）として導入したテーマを、別の参加者が簡単にトピック（BOKの第3階層）に変更できる。Wikipediaでは1回の編集で数百万ページ中の1ページしか更新できないのに対して、WikiBOKでは個人が全体に与える影響が大きすぎる。

### 1.2.1 SIBOK

第1の問題を解決する方法として、授業のシラバスの内容をBOK木の初期ノードとして導入する方法が提案され、青山学院大学社会情報学部の授業シラバスをもとにしたSocial Informatics Body of Knowledge（**SIBOK**）の構築が進められている[3]（つまり、社会情報学に関わる授業のシラバスから抽出した要素をBOK木の初期ノードとし、WikiBOKという手法で集団で構築されているBOKがSIBOKである）。2013年5月22日現在のSIBOKの上位2階層（エリアとユニット）は表1のようになっている。

### 1.2.2 科目の分類

第1の問題を解決する方法として、授業のシラバスを解析

表1 SIBOKの上位2階層。左側がエリア（11個）、右側がユニット（83個）である（2013年5月22日現在）。

人・組織・社会	情報社会 社会心理 組織とネットワーク
社会とメディア	広報広告 情報倫理 メディア論 ジャーナリズム コミュニケーション
ウェブと情報社会	ウェブと社会 ブログと社会 ネットコミュニティ 社会ネットワーク分析 インターネットテクノロジー Web入門 Webコンテンツ製作 Web2.0と知識社会
情報社会システム	情報社会の法 情報社会の政治 情報社会の教育 情報社会の経済 情報社会の行政
ウェブ社会とビジネス	情報化と企業 グローバル企業 ウェブと企業連携 ソーシャルビジネス 人工知能とビジネス ドットコムカンパニー ウェブ社会における起業
意思決定とリスク管理	大震災 社会保障制度 保険可能なリスク リスクマネジメント 意思決定に必要な要件 プロセスマネジメント概要
コンピュータと情報社会	コンピュータ インターネット プログラミング 情報管理と情報操作 コンピュータリテラシー 個人別プログラム開発 コンピュータサイエンス 情報倫理とセキュリティ オペレーティングシステム コンピュータの誕生と発達 インターネットの特性と課題 コンピュータグラフィックス ユーザインタフェースデザイン 人間中心設計とユーザビリティ 四則演算の論理回路による表現 ユーザインタフェースデザイン実習 コンピュータアーキテクチャと動作原理 ユニバーサルデザインとアクセシビリティ インタラクションの要素としてのコンピュータ コンピュータの基本構成とマイクロプロセッサ プログラミング概論、プログラム開発工程の説明 パーソナルコンピュータとクライアント/サーバコンピューティング HIIの進展
データマネジメントと分析	クラス分類 データ処理 データベース クラスタリング 情報蓄積と利用 テキストマイニング データの収集・整理 データの比較・整理 アソシエーションルール
地球環境と資源・エネルギー	原発問題 成長の限界 サステナブル社会 エネルギー問題と情報 情報化と持続可能な社会発展
社会情報抽出とハンドリング	確率 統計 仮説検定 回帰分析 標本調査 確率分布 社会調査 統計的推定 統計データ処理
情報システムとプロジェクトマネジメント	業務システム ソフトウェア工学 プロジェクトマネジメント

し、表2のように、授業が扱うテーマをユニット（BOKの第2階層）に、類似したテーマを持つ授業の集まりをエリア（BOKの第1階層）にする方法も提案されている[4]。しかしこの方法は、授業を分類した結果をBOKとしている点に問題がある。文献[1]で例示されているように、両者の関係は表3のように

表2 文献[4]におけるカリキュラムとBOKの関係性(○は授業でユニットが扱われていることを意味する)

	BOK～カリキュラム			
	エリア1		エリア2	
	授業a	授業b	授業c	授業d
ユニットA	○			○
ユニットB	○		○	○
ユニットC		○		
ユニットD		○	○	○
ユニットE		○		○

表3 文献[1]におけるカリキュラムとBOKの関係性(○は授業でユニットが扱われていることを意味する)

			カリキュラム			
			授業a	授業b	授業c	授業d
BOK	エリア1	ユニットA	○			○
		ユニットB	○		○	○
		ユニットC		○		
	エリア2	ユニットD		○	○	○
		ユニットE		○		○

表現されるべきである。そこでは、BOKのユニット全体を覆うように授業科目群が作られ、カリキュラムが構成されている。授業を分類してもBOKにはならない。カリキュラムとBOKは、教育内容の別々の軸への射影なのである。

### 1.2.3 ユニットのクラスタ分析

本稿では、科目ではなくユニットを、アルゴリズムで分類することによってBOKを構築する手法を提案する。これは、表3のようにBOKをとらえているため、「科目の分類」について前項で指摘した問題は生じない。さらに、先述の問題点2,3も、この方法では回避される。たとえば「統計学」の重要度は分野によって異なるが、1科目なのか1時間なのかといった、カリキュラム中での扱いで重要度はわかるため、問題2は回避される。また、BOK木の構築は参加者による編集ではなくアルゴリズムによって行われるため、問題3も回避される。

本手法はすでに、CC2001を構成する授業からCSBOKを再現する試みで用いられ、その有効性が確認[5]されている。

### 1.2.4 集約のメカニズム

集合知が機能するためには、集団の意見に**多様性**と**独立性**、**分散性**があり、そのような意見を**集約するメカニズム**が必要だと考えられている[6]。

本研究ではこの中でも特に集約のメカニズムに着目する。複数人の意見を集約して、1つの成果物を作ろうとする試みでは、以下のような集約のメカニズムが知られている。

**牛の体重やビンの中のジェリービーンズの数の予測** 複数人の予測を意見として入力し、単純平均という集約のメカニズムを活用すると、予測値という成果物が出力される(集団における個々人の推測の誤差が多様性によって相殺される[7])。

**Google PageRank**[8] ウェブページが持つリンク情報を意見として入力し、PageRankの計算という集約のメカニズムを活用すると、PageRankという成果物が出力される(よいページか

らリンクされているページがよいページになる)。

提案手法は、以下の2点において集合知の活用を試みている。

(1) シラバスという複数人によって独立に記述された情報を利用する。

(2) クラスタ分析の前処理(2.2節)として、用語を揃える作業を複数人で共同で行う。

第1の点について、シラバスを意見として入力し、BOKという成果物が出力されるためには、どのような集約のメカニズムがあればよいか。そのメカニズムとして、「クラスタ分析」を提案するのが本稿の趣旨である。

第2の点のためには別のメカニズムを考案するのが望ましいが、本稿ではSIBOKの成果を代用することにする(2.2節)。

## 2. 手 法

本節では、シラバスをもとに社会情報学のBOKを作成する手法を説明する。

### 2.1 シラバスの整理

青山学院大学社会情報学部の2012年度のカリキュラムにおいて、履修の要となる「コア」と呼ばれる38科目を利用する。シラバスには、授業のテーマが1回分ずつ計15個記述されているのが一般的だが、これらのテーマをBOKの「ユニット候補」として取り出す。

### 2.2 前 処 理

クラスタ分析の前に、ユニット候補を以下のように整理する。  
**用語の統一** 同一の概念がシラバス毎に違った用語で呼ばれているものを探し出し、それらを統一する。

**上位概念への変換** 社会情報学のBOKのユニットとしては細かすぎるとされる用語を探し出し、それらをより抽象度の高い用語に変換する。

ここでは、現在構築が進められているSIBOKの2013年5月22日現在のデータを活用することによって、これらの整理を行うこととする。SIBOKはシラバスから抽出したユニット候補をBOK木の初期ノードとしているため、その構築作業の中に、ここで述べている前処理が含まれているからである。ただし、SIBOKを活用するのは用語の統一と上位概念への変換、つまり個々のユニット候補の直近の情報のみであり、「エリアの種類」といったBOK全体についての情報は参照しない。

### 2.3 ユニットのクラスタ分析

クラスタ分析には、階層的クラスタ分析と非階層的クラスタがあるが、BOK構築には非階層的クラスタ分析のほうが有望だという予備実験結果[5]があったため、ここでは階層的クラスタ分析のみを用いる。

クラスタ分析の詳細を、表3を例に説明する。

表3のようなユニットと授業の関係は、式(1)の左辺のような行列で表現できる。この行列の各列を規格化する。つまり、各列の要素の2乗和が1になるように、列毎に定数をかける(式(1)の右辺)。この操作は、授業が結果に与える影響の大きさを揃えるために行う(予備実験によって、この操作を行った方が現実的なBOKが得られることがわかっている)。こうして

できる行列の各行を、ユニットを表現する特徴ベクトルとみなし、それらをクラスタ分析する。

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1/\sqrt{2} & 0 & 0 & 1/2 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} & 1/2 \\ 0 & 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{3} & 1/\sqrt{2} & 1/2 \\ 0 & 1/\sqrt{3} & 0 & 1/2 \end{pmatrix} \quad (1)$$

階層的クラスタ分析の実行時には、要素同士の非類似度の指標とクラスタの連結法、生成させるクラスタ数を任意に決められる。それぞれの選択肢を以下に挙げる<sup>(注1)</sup>。

### 2.3.1 非類似度の指標

2つのベクトル（本稿ではユニットの特徴ベクトル） $u, v$ の非類似度の指標には次のようなものがある。

- ユークリッド距離： $|u - v|$
- 平方ユークリッド距離： $|u - v|^2$
- マンハッタン距離： $\sum |u_i - v_i|$
- チェビシェフ距離： $\max(|u_i - v_i|)$
- キャンベラ距離： $\sum |u_i - v_i| / (|u_i| + |v_i|)$
- 余弦距離： $1 - u \cdot v / (|u||v|)$
- 相関距離： $1 - \sum (u_i - \bar{u})(v_i - \bar{v}) / \sqrt{\sum (u_i - \bar{u})^2 \sum (v_i - \bar{v})^2}$
- プレイ・カーティス距離： $\sum |u_i - v_i| / \sum |u_i + v_i|$

ここで、 $u_i$ は $u$ の第 $i$ 成分、 $\bar{u} = \sum u_i / (u$ の要素数)である。

### 2.3.2 クラスタの連結法

階層的クラスタ分析で用いるクラスタの連結法には、次のようなものがある。

- 最短距離法
- 群平均法
- 最長距離法
- 重み付き群平均法
- クラスタの重心からの距離
- クラスタのメジアンからの距離
- Wardの最小分散法

### 2.3.3 クラスタ数

ユニットをクラスタ分析することによってできるクラスタをBOKのエリアと見なす。SIBOKの2013年5月22日現在のエリア数が11であることを考慮して、クラスタ分析で生成するクラスタの数は10, 11, 12のいずれかとする。

## 2.4 クラスタ分析結果の評価

非類似度の指標が8種類、クラスタの連結法が7種類、生成させるクラスタ数が3種類あるため、全部で $8 \times 7 \times 3 = 168$ 通りのクラスタ分析結果が得られる。これらの結果は、以下の2つの方法で評価する。

### 2.4.1 教員へのアンケート

青山学院大学社会情報学部「社会情報学」を表現するものとして構築されたBOKが、その実態を反映したものになっているかどうかを、複数名の教員へのアンケートによって評価する。アンケートは、BOKが実態を反映しているかどうかを5段階

階で評価するものとする。評価結果を集計することによって、この問題にふさわしいクラスタ分析方法を探る。

### 2.4.2 SIBOK との比較

本手法で構築したBOKをSIBOK（表1）と比較することでも、本手法の有効性を議論できる。

BOKの比較には、エリアの構成要素であるユニットのペアを考える。例えば、表1のエリア「人・組織・社会」のユニットから、{情報社会, 社会心理}・{社会心理, 組織とネットワーク}・{組織とネットワーク, 情報社会}という3個のペアを作れる。このようなペアについての、以下に挙げる指標を用いて、2つのBOKを比較する。これらの指標はCC2001を構成する授業からCSBOKを再現する試みで用いられたものである[5]。

**再現率** SIBOKから作られるペアで、本手法で構築したBOKから作られるペアの集合に含まれるものの割合

**精度** 本手法で構築したBOKから作られるペアで、SIBOKから作られるペアの集合に含まれるものの割合

**F値** 再現率と精度の調和平均

## 3. 結 果

### 3.1 シラバスの整理

青山学院大学社会情報学部の2012年度のカリキュラムから、ユニット候補として取り出されたテーマは365個であった。

### 3.2 前 処 理

用語を統一した例には次のようなものがあつた。

- 「仮説検定」と「統計的仮説検定」を「仮説検定」に統一
- 「インターネットの技術」と「インターネットテクノロジー」を「インターネットテクノロジー」に統一

上位概念への変換例には次のようなものがあつた。

- 授業「システム分析・設計」で扱われる4つのユニット候補：「基礎要求分析」、「モデリング」、「要求分析」、「システムライフサイクルとシステム開発プロセス」を「ソフトウェア工学」に変換
- 授業「ヒューマンインタフェース」で扱われる1つのユニット候補：「システム操作の人間特性」を「ソフトウェア工学」に変換
- 授業「システム分析・設計応用」で扱われる2つのユニット候補：「ステークホルダー分析、ユースケース」と「改善案の検討」を「ソフトウェア工学」に変換

このような作業の結果、最終的に利用するユニットは62個になった（表6）。SIBOKにはこれより多い83個のユニットがあるが（表1）、表1にあつて表6にはない21個は、授業とは結びつけられなかったユニットである。

### 3.3 ユニットのクラスタ分析

2.4節で述べた168通りのクラスタ分析の結果には、ユニットが1つしかないようなクラスタを持つものが142個あつた。それらはBOKとしてはふさわしくないと考え、不採用とした。平方ユークリッド距離を非類似度の指標にしたものはすべてこれに該当したため、以下の結果の中に平方ユークリッド距離は現れない。

(注1)：ここで挙げる非類似度の指標とクラスタ連結法は、本研究で利用したソフトウェア（Mathematica Ver. 9.0）に組み込まれているものすべてである。

表4 クラスタ分析結果に対する教員（3名）の評価（空白部分は不採用だったもの）

	最長距離法			Ward の最小分散法		
	10	11	12	10	11	12
プレイ・カーティス	2,2,1	2,2,1		5,5,5	4,5,5	3,3,4
キャンペラ				3,4,2	3,3,2	3,3,3
チェビシェフ				2,1,2	2,2,2	2,2,2
相関	2,2,1	2,2,1	2,2,1	4,5,5	4,4,3	3,4,4
余弦	2,2,1			5,5,5	4,5,5	4,4,4
ユークリッド				2,2,3	4,2,3	
マンハッタン				3,4,3	2,3,2	2,3,3

表5 アンケート結果（表4）の分散分析結果

	F 値	Pr(>F)
非類似度の指標	11.932	2.24e-08
クラスタ連結法	222.181	< 2e-16
クラスタ数	3.598	0.0344
非類似度の指標：クラスタ連結法	1.623	0.2072
非類似度の指標：クラスタ数	2.100	0.0368
クラスタ連結法：クラスタ数	1.490	0.2349
3 因子	0.542	0.4650

### 3.4 クラスタ分析結果の評価

#### 3.4.1 教員へのアンケート

青山学院大学社会情報学部の教員3名が、クラスタ結果を評価した結果を表4に掲載する。3つの数字が3名の教員の評価点である。各教員は、それぞれの結果を5点満点で採点した。評価が最も高かったのは、非類似度の指標としてプレイ・カーティス距離あるいは余弦距離を、クラスタの連結法としてWardの最小分散法を用い、10個のクラスタを生成させた場合の結果である（表中の太字）。2つの結果は、エリアとユニットの順番が違っただけで、ユニットの分類としては同じものであった（プレイ・カーティス距離を使った結果を表6に掲載する）<sup>(注2)</sup>。

表4を分散分析した結果を表5に掲載する。評価値への影響が有意なのは、非類似度の指標（単独）とクラスタ連結法（単独）、クラスタ数（単独）、非類似度の指標とクラスタ数（交互作用）であった（有意水準は5%）。この時点で、クラスタ連結法はWardの最小分散法を使うのがよいことがわかる。また、Tukey HSDで分析すると（結果の詳細は割愛）、クラスタ数は12より10のほうがよいことがわかる（P値は0.033）。

クラスタ連結法をWardの最小分散法に、クラスタ数を10または11に限定した結果を分析（分散分析およびTukey HSD）すると（結果の詳細は割愛）、最終的には以下のような結論になった（有意水準は5%）。

**非類似度の指標** プレイ・カーティス距離と相関距離、余弦距離がよい

**クラスタ連結法** Wardの最小分散法がよい

**クラスタ数** クラスタ数は10あるいは11にするのがよい

アンケートに答えた教員からは、「表6のエリア4とエリア

5は連結してもよいと思われるが、クラスタ数を1減らして9にすれば連結されるか」という疑問が呈されたが、クラスタ分析結果を詳しく調べると、次に連結されるのはエリア2とエリア3であり、エリア4とエリア5ではないことがわかった

教員からはまた、「プログラミング」と「ログと社会」が同じエリアになる理由も問われた。クラスタ分析結果を詳しく調べると、「組織とネットワーク」から「コンピュータリテラシ」までは、共通の授業がなく最後まで連結しなかったユニット群であった。これらが連結してできたクラスタと、授業「社会情報学体験演習」という共通の授業を持つ「プログラミング」と「データ処理」が連結してできたクラスタによって、エリア3はできていた。

#### 3.4.2 SIBOK との比較

本手法で構築したBOKとSIBOK（表1）を、2.4.2項で導入した指標（再現率と精度、F値）を用いて比較した結果を表7に掲載する。教員評価と各指標との相関係数は、再現率が-0.28、精度が0.71、F値が0.18であった<sup>(注3)</sup>。

## 4. 考 察

本研究は、シラバスという大学教員が自分の専門分野について作成した小さな知識体系を「集約」し、学問分野を形作る大きな知識体系にする研究である。集約のためのメカニズムとして、「クラスタ分析」を利用することを提案し、実際に実行する際の具体的な指針が得られた。

非類似度の指標は、プレイ・カーティス距離と相関距離、余弦距離を使うのがよいという結果になった。これは、CC2001を構成する授業からCSBOKを再現する試みの結果[5]と矛盾しない。これらの結果は、本手法には幾何学的な意味づけの弱い非類似度の指標が向いているということを示唆しているが、結論を下すためにはさらなる研究が必要であろう。

本手法の結果とSIBOKの比較において、2.4.2項で導入した指標（再現率と精度、F値）の中で、教員の評価との相関が最もよかったのは精度であった。これは、ここで用いた再現率が、2つのBOKの類似度のよい指標ではないということを示唆している。この原因は、大きなクラスタがありさえすれば、ユニットのペアの数が増え、再現率が高くなることだと思われる。

システムの改善点として、3.4.1項で紹介したような利用者からの疑問にインタラクティブに対応できるシステムがあると、作業がしやすくなると思われる。

見出されたクラスタ（エリア）に名前を付けるためには、何か別の方法が必要であろう。

## 5. おわりに

集合知による知識体系構築のための意見集約手法として、クラスタ分析を用いることを提案した。実例として、「社会情報学」を学ぶ教育機関のシラバスを意見として抽出し、それをクラスタ分析することによって、「社会情報学」の知識体系の構築

(注2)：非類似度の指標として相関距離を使った結果も、ユニットの分類としては同じものだったが、1人の教員は順番の違いで点数を変えたと思われる。

(注3)：本手法の結果とSIBOKでは、3.2節で述べたようにユニット数に差があるが、ここでの比較には、両方に存在するユニットのみを用いている。

表6 教員からの評価が最も高かった結果の1つ（非類似度の指標：ブレイ・カーティス距離、クラスタ連結法：Wardの最小分散法）

エリア1	統計 仮説検定 統計データ処理 回帰分析 標本調査 確率分布 統計的推定 データの収集・整理
エリア2	プロジェクトマネジメント リスクマネジメント 社会保障制度 保険可能なリスク 意思決定に必要な要件 プロセスマネジメント概要
エリア3	業務システム ソフトウェア工学 情報社会の政治 情報社会の経済 組織とネットワーク 情報社会の行政 コミュニケーション インターネットの特性と課題 ブログと社会 コンピュータリテラシ データ処理 プログラミング 情報蓄積と利用 情報管理と情報操作
エリア4	大震災 原発問題 成長の限界
エリア5	エネルギー問題と情報 情報化と持続可能な社会発展 情報社会 情報化と企業
エリア6	情報社会の法 インターネット インターネットテクノロジー ネットコミュニティ
エリア7	クラス分類 クラスターリング テキストマイニング アソシエーションルール
エリア8	Web2.0と知識社会 ウェブと社会 人工知能とビジネス 情報倫理とセキュリティ コンピュータグラフィックス
エリア9	インタラクションの要素としてのコンピュータ ユニバーサルデザインとアクセシビリティ HIの進展 ユーザインタフェースデザイン実習 ユーザインタフェースデザイン 人間中心設計とユーザビリティ
エリア10	パーソナルコンピュータとクライアント/サーバコンピューティング データベース コンピュータの基本構成とマイクロプロセッサ コンピュータサイエンス オペレーティングシステム コンピュータの誕生と発達 四則演算の論理回路による実現 コンピュータアーキテクチャと動作原理

を試みた。得られた結果を教員が評価したり、人手で作成されたBOKと比較したりすることによって、提案手法を用いる際の具体的な指針が得られた。本手法の利用環境を整備し、さまざまな分野で応用することによって、集合知のための意見集約手法の一つとしての有効性を高められることが期待される。

表7 本研究の結果とSIBOKの比較（連結法のCは最長距離法、WはWardの最小分散法）

クラスタ分析設定			比較結果			教員 評価
非類似度の指標	連結法	エリア数	再現率	精度	F値	
ブレイ・カーティス	C	10	0.288	0.205	0.24	5
ブレイ・カーティス	C	11	0.27	0.248	0.259	5
ブレイ・カーティス	W	10	0.317	0.428	0.364	15
ブレイ・カーティス	W	11	0.302	0.462	0.366	14
ブレイ・カーティス	W	12	0.26	0.435	0.325	10
キャンベラ	W	10	0.292	0.279	0.285	9
キャンベラ	W	11	0.274	0.377	0.318	8
キャンベラ	W	12	0.253	0.364	0.298	9
チェビシェフ	W	10	0.256	0.33	0.289	5
チェビシェフ	W	11	0.235	0.314	0.269	6
チェビシェフ	W	12	0.228	0.311	0.263	6
相関	C	10	0.616	0.35	0.446	5
相関	C	11	0.548	0.395	0.459	5
相関	C	12	0.544	0.396	0.459	5
相関	W	10	0.317	0.428	0.364	14
相関	W	11	0.274	0.399	0.325	11
相関	W	12	0.26	0.432	0.324	11
余弦	C	10	0.327	0.22	0.263	5
余弦	W	10	0.317	0.428	0.364	15
余弦	W	11	0.302	0.462	0.366	14
余弦	W	12	0.26	0.432	0.324	12
ユークリッド	W	10	0.281	0.328	0.303	7
ユークリッド	W	11	0.263	0.351	0.301	9
マンハッタン	W	10	0.345	0.386	0.365	10
マンハッタン	W	11	0.302	0.411	0.348	7
マンハッタン	W	12	0.281	0.399	0.33	8

## 謝 辞

クラスタ分析結果についてのアンケートにご協力いただいた、青山学院大学の福田亘孝教授に感謝いたします。

## 文 献

- [1] The Joint Task Force on Computing Curricula IEEE Computer Society and Association for Computing Machinery, “Computing curricula 2001 computer science,” 2001.
- [2] 増永良文, 石田博之, 伊藤一成, 伊藤守, 清水康司, 莊司慶行, 高橋徹, 千葉正喜, 長田博泰, 福田亘孝, 正村俊之, 森田武史, 矢吹太朗, “知識体系(BOK) 創成支援システム WikiBOK の研究・開発,” 第3回ソーシャルコンピューティングシンポジウム講演論文日本データベース学会, pp.67–72 2012.
- [3] 増永良文, 石田博之, 伊藤一成, 伊藤守, 清水康司, 莊司慶行, 高橋徹, 千葉正喜, 長田博泰, 福田亘孝, 正村俊之, 森田武史, 矢吹太朗, “WikiBOK を用いた社会情報学の知識体系構築実験,” 日本データベース学会論文誌, vol.12, no.1, 2013, (受理).
- [4] 川端智久, 白井靖人, “授業内容に基づく知識体系の構築,” 第24回セマンティックウェブとオントロジー研究会人工知能学会, pp.12–1–12–7 2011. SIG-SWO-A1101-12.
- [5] 矢吹太朗, 増永良文, 森田武史, 石田博之, “知識体系のエリア自動抽出のためのユニット分類手法,” 第5回データ工学と情報マネジメントに関するフォーラム (DEIM2013), 2013.
- [6] J. Surowiecki, The Wisdom of Crowds, Anchor, 2005. 小高尚子訳, 「みんなの意見」は案外正しい, 角川書店, 2006.
- [7] S.E. Page, The Difference, Princeton Univ. Press, 2007. 水谷淳訳, 「多様な意見」はなぜ正しいのか, 日経 BP 社, 2009.
- [8] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” Proceedings of the seventh international conference on World Wide Web 7, pp.107–117, WWW7, 1998.