

平成 20 年度
学士学位論文

球面 SOM を用いたシラバス可視化 システムの構築

Development of Syllabus Visualization System using
Spherical Self-Organizing Maps

1090366 浜田 洋

指導教員 妻鳥 貴彦

2009 年 3 月 5 日

高知工科大学 情報システム工学科

要 旨

球面 SOM を用いたシラバス可視化システムの構築

浜田 洋

近年，多くの大学でシラバスが導入されている．シラバスは，大学で行われる講義の概要を示した冊子であり，主に学生が履修計画を立てる際の参考に使っている．シラバスには科目間の関係を示す情報も記載されており，その情報を学生が理解することは体系的な履修計画を立てるために重要である．しかし，シラバスには科目間の関係が読み取りづらいという問題点がある．そこでシラバスを可視化し，科目間の関係を読み取りやすくする研究が行われてきた．

先行研究で SOM(Self-Organizing Maps) を用いたシラバス可視化システムが提案された．SOM は，マップ上のノードにデータを学習させることで，データを類似度によって分類して表示する．先行研究では，SOM の学習結果を平面のマップを用いて可視化していた．しかし，SOM を用いて平面に可視化すると，マップの端において可視化結果に歪みが生じるという問題がある．

本研究では，先行研究での問題を解決するために，球面 SOM を用いたシラバス可視化システムを構築した．また，システムの有効性を確認するために，シラバス作成者である教員とシラバスの主な利用者である学生を対象に評価アンケートを実施した．評価の結果，本システムは先行研究よりも科目間の関係を読み取りやすくなっているが，操作性や表示方法に改善の余地があることがわかった．

キーワード シラバス，シラバス可視化，SOM，球面 SOM

Abstract

Development of Syllabus Visualization System using Spherical Self-Organizing Maps

Hiroshi HAMADA

In recent years, syllabi have been introduced into most of universities. Syllabus is a booklet that shows course outline, and it is used for making students' study plan. Syllabus has information to show relations between subjects. It is important that students understand the information of syllabus for the purpose of making systematic study plan. However, syllabi have problem what is that syllabi are difficult to understand the subjects relations. To solve the problem, syllabus visualization was proposed to easy to understand the subjects relations.

A study of syllabus visualization using Self-Organizing Maps(SOM) was proposed as a past study. SOM makes node on a map learning data, and data is classified by their similarity. In the past study, the result of SOM was visualized by 2D map. However, the method of the past study has a problem that the result of visualizing become distorted at edge of map.

In this paper, to solve the problem in the past study, we proposed syllabus visualizing system using spherical SOM. Moreover, to evaluate the validity of the system, we send out questionnaires to teachers and students. As a result, we confirmed that the system is easier to understand the subjects relations than the past study.

key words Syllabus, Syllabus Visualization, SOM, Spherical SOM

目次

第 1 章	はじめに	1
第 2 章	研究背景	2
2.1	シラバスとは	2
2.2	シラバスの問題点	2
2.3	問題の解決方法	4
2.3.1	SOM を用いたシラバスの可視化	4
2.3.2	SOM を平面にマッピングする場合の問題点	5
第 3 章	シラバス可視化システムの設計	7
3.1	SOM のアルゴリズム	7
3.2	球面 SOM	10
3.3	測地線ドーム	11
3.4	キーワードの分類	12
3.5	科目情報の表示方法	14
3.6	可視化結果の表示方法	14
第 4 章	シラバス可視化システムの実装	15
4.1	システムの構成	15
4.2	システムの概要	15
4.3	球面 SOM の実装	15
4.4	単一科目の情報の表示	16
4.5	ソサエティの割合によるベクトル間の距離計算	18
4.6	科目検索	19
4.7	球の裏側と展開図の表示	19

目次

第 5 章	システムの評価	23
5.1	評価方法	23
5.2	評価結果	26
5.3	評価結果の考察	28
第 6 章	まとめ	30
謝辞		31
参考文献		32

目次

2.1	高知工科大学情報システム工学科のプレレキジット表	3
2.2	先行研究のシステムの実行画面	5
3.1	SOM の概念図	8
3.2	近傍関数で学習されるノードの遷移	10
3.3	初期状態	12
3.4	1 回繰り返した状態	12
3.5	2 回繰り返した状態	12
4.1	実行画面	16
4.2	「集積回路システム」のソサエティの割合	17
4.3	「論理回路理論」のソサエティの割合	18
4.4	科目検索のウィンドウ	20
4.5	「論理回路理論」の検索結果	21
4.6	裏側の表示画面	21
4.7	展開図の表示画面	22
5.1	「問 1 , 問 2」のウィンドウ	24
5.2	「論理回路理論」に注目した画面	24
5.3	「画像処理」に注目した画面	24
5.4	「問 3 . 裏側」のウィンドウ	25
5.5	「問 3 . 展開図」のウィンドウ	25
5.6	「問 4」のウィンドウ	25

表目次

3.1 シラバス項目の語彙	13
4.1 ソサエティと色の対応	17
5.1 評価アンケートの質問項目とその回答群	26
5.2 問 1 , 問 2 に対する回答	26
5.3 問 3 に対する回答	27
5.4 問 4 に対する回答	27

第 1 章

はじめに

現在，多くの大学でシラバスが導入されている．シラバスは，大学で開講される科目の概要を記した冊子である．シラバスの主な利用者は学生であり，学生は履修計画を立てる際や予習の参考にするためにシラバスを参照する．近年では，シラバスを冊子ではなく電子媒体として Web 上で公開している大学も多くなっている．

しかし，現在のシラバスには科目間の関係が理解しづらいという問題点がある．科目間の関係には，カリキュラム上での前後関係がある履修前提科目と，履修前提科目ではないが同じ分野を学習する科目とがある．前者の関係は，シラバスに明記されている場合が多いが，後者の関係は明記されていない場合が多い．後者の関係を理解することは，学生にとって体系的な履修計画を立てる上で重要であるが，学生がその関係を理解するためには学生がすべての科目のシラバスから読み取らなければならない．この作業は履修計画を立てる学生にとって負担となる．

シラバスを利用者にとって分かりやすくするために，現在様々な手法でシラバス情報の可視化が提案されている．先行研究 [1] では，SOM(Self-Organizing Maps) を用いたシラバスの可視化手法が提案された．これは，シラバス中のキーワードを抽出し，そのキーワードを特徴量として SOM を適用することで，類似する科目を平面にマッピングするというものである．しかし，SOM には平面にマップを作成するとマップの端において可視化結果に歪みが生じるという問題がある．

本研究では，SOM を平面にマッピングすることによる可視化結果の歪みを抑制するために，SOM を球面上にマッピングする球面 SOM を用いることでシラバスを可視化するシステムを提案する．

第 2 章

研究背景

現在多くの大学でシラバスが導入されているが、シラバスには利用者である学生にとって科目間の関係が理解しづらいという問題点がある。本章では、シラバスの解説を行い、シラバスの問題点とその解決方法を述べる。

2.1 シラバスとは

シラバスとは、大学で開講される科目の情報が文章として科目ごとにまとめられた冊子である。シラバスに記載されている科目の情報は、その科目で学習する内容の概要やその科目を履修する前に備えておくべき知識、または事前に履修することが前提となっている履修前提科目などがある。

シラバスに記載されている情報は、学生が履修計画を立てる際の重要な参考資料となる。学生は履修計画を立てる際にシラバスを参照し、その科目を履修する上で必要な知識を備えておくことで、講義の内容をより深く理解することができる。

2.2 シラバスの問題点

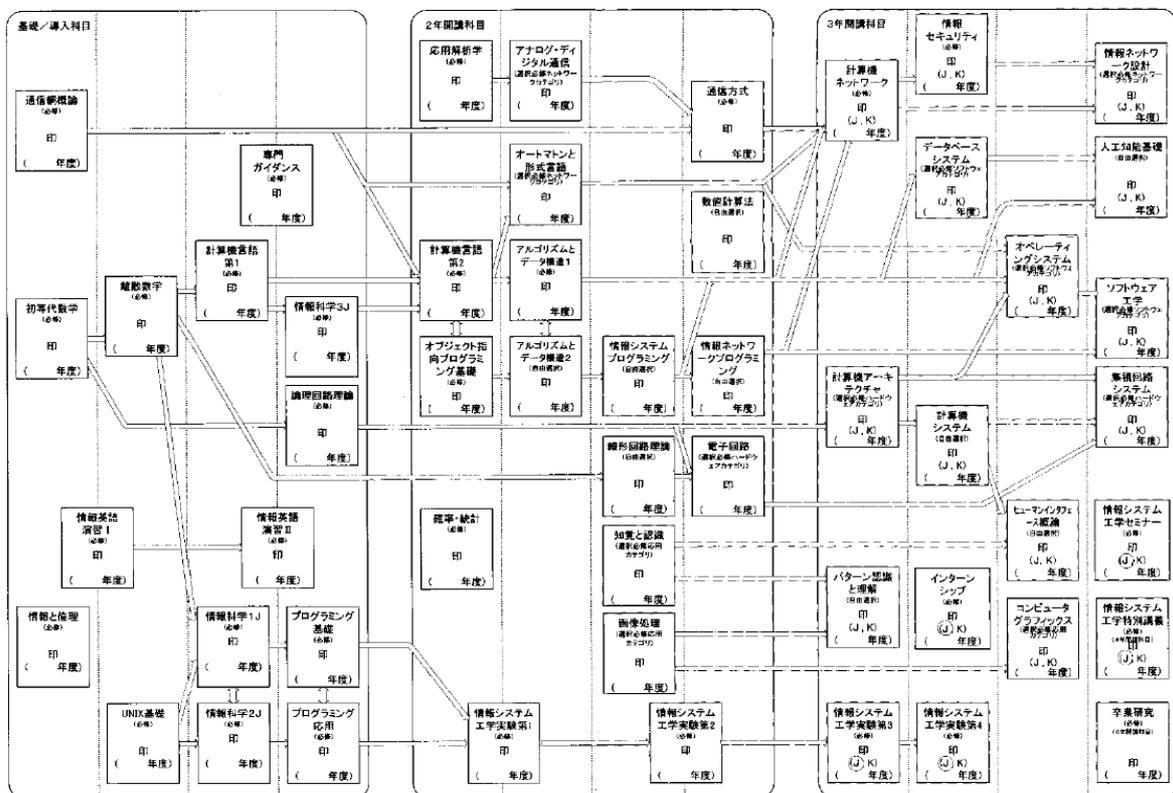
学生は、現在の単位取得状況や自身の関心がある分野などを考慮して、体系的に履修計画を立てることが望ましい。体系的に履修計画を立てるためには、学生は科目ごとの学習内容を理解するとともに、科目間の関係を理解する必要がある。

シラバスには科目間の関係を表す情報として履修前提科目がある。これはカリキュラムであらかじめ定められている履修の順序であり、多くの場合は基礎科目から応用・発展科目と

2.2 シラバスの問題点

続いている。履修前提科目はシラバス中で項目を設けて明記されている場合が多く、そこに書かれている科目をたどることで科目間の関係を知ることができる。履修前提科目はカリキュラム上での必須条件であり、学生が履修し忘れた場合に他の科目の履修を妨げることになるため、履修前提科目の項目をたどるよりも簡単なグラフで図示されている場合がある。履修前提科目を図示した例として、高知工科大学情報システム工学科のプレレキジット表を図 2.1 に示す。この図では、科目を開講時期によってまとめ、前後関係のある科目は矢印でつなぐことで関係を表している。

情報ネットワークシステムコース 専門科目のプレレキジット



注)このプレレキジット表に記された履修前提科目(詳)の単位を全く取得していない場合、当該科目の単位は取得できません。

図 2.1 高知工科大学情報システム工学科のプレレキジット表

もうひとつの科目間の関係として、ある科目の学習内容がその他の科目の学習内容に含まれるという関係がある。この関係は、ある科目とその履修前提科目の間にも成立していることが多い。しかし、履修前提科目に指定された科目以外でもこの関係が成立している場合が

2.3 問題の解決方法

ある．例えば，高知工科大学の「数学 8」と「応用解析学」，「アナログ・デジタル通信」の間にもこの関係は成立している．「応用解析学」は「アナログ・デジタル通信」の履修前提科目となっており，「アナログ・デジタル通信」で必要になるフーリエ級数とフーリエ変換を学習する．同じくフーリエ級数とフーリエ変換を学習する科目として「数学 8」があり，「数学 8」のシラバスにもそのように記載されている．しかしながら，「アナログ・デジタル通信」のシラバスに履修前提科目として明記されているのは「応用解析学」のみであり，「数学 8」との関係を読み取ることは容易ではない．

このような関係を理解することは学生にとってある分野をより深く学習できるため有益であると考えられる．しかし，現在のシラバスで学生がこの関係を読み取るためには，学生はシラバスをすべて読み，同じ分野を学習するものがないかを調べなければならない．この作業はこれから履修計画を立てる学生にとって困難であり，有益な情報を取得できない可能性がある．

2.3 問題の解決方法

2.3.1 SOM を用いたシラバスの可視化

先行研究 [1] で，シラバスの問題点を解決するために，SOM (Self-Organizing Maps) [2] を用いたシラバス可視化のシステムが提案された．このシステムでは，シラバスの文章中から科目の学習内容を表しているキーワードを抽出し，その出現頻度を特徴量として科目を分類する．

シラバス中に出現するキーワードは種類が多く，科目によって使用しているキーワードが異なる．そのため，キーワードの出現頻度を特徴量とするとそれぞれに異なる次元をもった高次元のデータとなり，単純な比較ができない．そこで先行研究では，このようなデータの類似性を明らかにするのに適した手法として，SOM を用いたシラバスの可視化を行った．システムの実行画面を図 2.2 に示す．

図 2.2 のシステムでは，同じ矩形範囲にある科目同士が類似した内容を学習すると判断さ

2.3 問題の解決方法

Self-Organizing Map [10000/10000(終了)]					
応用解析学	情報理論基礎	計算機アーキテクチャ情報 オートマトンと形式言語		オブジェクト指向プログラミング 情報システム工学実験第1 計算機言語第2	ヒューマンインターフェース概論
パターン認識・理解	集積回路システム 電子回路情報 線形回路理論 専門ガイダンス				
アナログ・デジタル通信		知覚と認識 情報代数	アルゴリズムとデータ構造2 アルゴリズムとデータ構造1	オペレーティングシステム	
情報セキュリティ		画像処理 データベースシステム 離散数学	情報ネットワークプログラミング 情報システムプログラミング 情報システム工学実験第2	通信方式 通信網概論	
論理回路理論		人工知能基礎	ソフトウェア工学	計算機ネットワーク	
		コンピュータグラフィックス	数値計算法 感覚・神経生物学	計算機言語第1	情報ネットワーク設計

図 2.2 先行研究のシステムの実行画面

れている。また、科目が集中している矩形範囲に限定して、その中でさらに分類することで細かい分類を示している。

2.3.2 SOM を平面にマッピングする場合の問題点

SOM のアルゴリズムでは、あるノードに学習させる際にその近傍ノードにも学習の影響が出る。言い換えれば、あるノードはその周辺のノードの学習結果に影響を受けながら学習を進めるということである。SOM の出力層となるマップを平面としたとき、マップの端にあるノードでは中央付近のノードに比べて近傍ノードが少なくなる。このためマップの端のノードでは影響を受けるノードが少なくなり、学習結果に歪みが生じてしまう。

2.3 問題の解決方法

このような問題を解決するために本研究では球面 SOM を用いてシラバス可視化のシステムを構築する。球面 SOM とは、基本的なアルゴリズムは SOM と変わらず、出力先のノードが球面上に配置されたものである。これによりマップの端が存在しなくなるので、学習の歪みを抑制することができる。

第 3 章

シラバス可視化システムの設計

シラバスの問題点を解決するために，シラバスの可視化を提案する．本章では，SOM を平面にマッピングして可視化する場合の問題点を解決するために，球面 SOM を用いたシラバス可視化システムの設計について述べる．シラバスからの情報抽出と抽出したキーワードをソサエティに分類する機能は先行研究 [1] のシステムを利用する．

3.1 SOM のアルゴリズム

SOM は Teuvo Kohonen によって提案されたニューラルネットワークを用いたデータ可視化手法の一種である [2]．これは中間層ニューロンを用いない教師なし競合学習モデルである．SOM は，多次元ベクトルで表現される多数のデータからある傾向があるデータを発見したり，クラス分けを行うために用いられる．

SOM は入力層と出力層の 2 層から構成される．入力層はノードが並んだ行

$$N_{\text{in}} = (m_1 \quad m_2 \quad \cdots \quad m_L)$$

で表現され，出力層はノードが格子状に並んだ行列

$$N_{\text{out}} = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1N} \\ m_{21} & m_{22} & \cdots & m_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{M1} & m_{M2} & \cdots & m_{MN} \end{pmatrix}$$

で表される．入力層，出力層のノード m_k はそれぞれ n 次の重みベクトル

$$w_k = (w_{k1}, w_{k2}, \cdots, w_{kn})$$

3.1 SOM のアルゴリズム

を持つ．入力層ノードの持つ重みにより，出力層ノードの重みへの学習が繰り返し行われる．図 3.1 に SOM の概念図を示す．

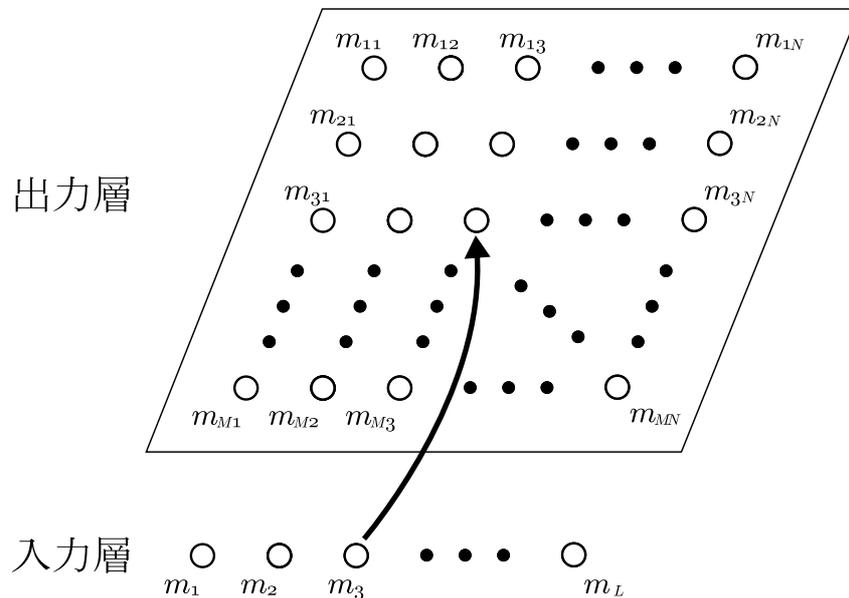


図 3.1 SOM の概念図

SOM における学習は，いくつかの手法が提案されているが，本研究においては以下のアルゴリズムで示す手法を用いる．このアルゴリズムに対して，定数値として学習繰り返し回数 λ および初期学習半径 R ，出力層の縦横のサイズ M および N が与えられる．一般に λ が 1000 回以上， R は出力層のサイズの 2 分の 1 程度となる．このマップ上において，入力データは出力層ノードの $M \times N$ 個のクラスにクラス分けされる．

ステップ 1. 初期化

入力層 N_{in} の各ノードが持つ重みを入力データを用いて初期化し，出力層 N_{out} の各ノードが持つ重みをランダムに初期化する．

ステップ 2. サンプルの選択

学習のサンプルとなるノードを入力層の先頭から順に選択する．ここで選択されたサンプルを m ， m が持つ重みを x とする．

ステップ 3. 勝者ノードの選択

3.1 SOM のアルゴリズム

まずサンプルの重み v と、出力層の各ノード m_{ij} の重み w_{ij} との距離をそれぞれ計算する。2つのベクトルの距離の計算には一般にユークリッド距離が用いられ、

$$\text{dist}(v, v^*) = \|v - v^*\|^2 = \sqrt{\sum_{i=1}^n (v_i - v_i^*)^2} \quad (3.1)$$

と表される。ここでは $\text{dist}(x, w_{ij})$ が最小の値をとる、サンプルと重みが最も良く似たノード m_{ij} が勝者ノードとなる。

ステップ 4. 重みの学習

出力層上で、勝者ノード m の近傍にあるノード m_{xy} に対し、サンプルの重み x を学習させ、その重み w_{xy} を更新する。学習後の重みの値は (3.2) 式で求められる。ここで、現在の学習回数を $0 \leq t \leq \lambda - 1$ とし、 $v^{(t)}$ は v の t 回の学習時点での値を意味する。

$$w_{xy}^{(t+1)} = (1 - T) \cdot w_{xy}^{(t)} + T \cdot x \quad (3.2)$$

(3.2) 式中の T は、学習変数 $\alpha(t)$ と近傍関数 $h(m, m_{xy})$ で求められ、(3.3) 式で表される。

$$T = \alpha(t) \cdot h(m, m_{xy}) \quad (3.3)$$

式中の $\alpha(t)$ は、

$$\alpha(t) = 1 - \frac{t}{\lambda} \quad (3.4)$$

と表され、学習対象の重み $w_{xy}^{(t)}$ がどの程度サンプルの重み v を学習するかを決定する。 $\alpha(t)$ は、学習初期は 0 に近く、学習が進むにつれ 1 に近づいていく性質を持つ。そのため、(3.2) 式がとる値は、初期の学習ではサンプルの重み x から学習量が大きく、学習が進むにつれその影響が小さくなっていく。

また $h(m, m_{xy})$ は一般に

$$h(m, m^*) = \exp\left(\frac{\text{dist}(m, m^*)}{\sigma^2}\right) = \exp\left(\frac{-\|m - m^*\|^2}{(\alpha(t) \cdot R)^2}\right) \quad (3.5)$$

で示すガウス関数で求められ、この関数は勝者ノードと学習対象ノードの距離に応じた学習量を決定する。 $h(m, m_{xy})$ のとる値は、学習対象ノード m_{xy} が勝者ノード m に近いほど 1 に近づき、遠くなるほど 0 に近づく。また、 σ の値域は $0 < \sigma < R$ であり、学習が進むに

3.2 球面 SOM

つれ、0 に近づく。これらの性質から、(3.2) 式で学習が行われるノードは、学習の初期では勝者ノード m を中心とした初期学習半径 R に近い範囲となり、学習が進むにつれその範囲が狭くなっていく。(3.5) の値と、学習が行われるノードの遷移を図 3.2 に示す。円の中の色の濃淡は $h(m, m^*)$ の取る値を示し、これが黒に近いほど 1 に近いことを示す。また、点は出力層のノードを示し、円の中央の点が勝者ノード、白い点が学習対象となるノードを示している。

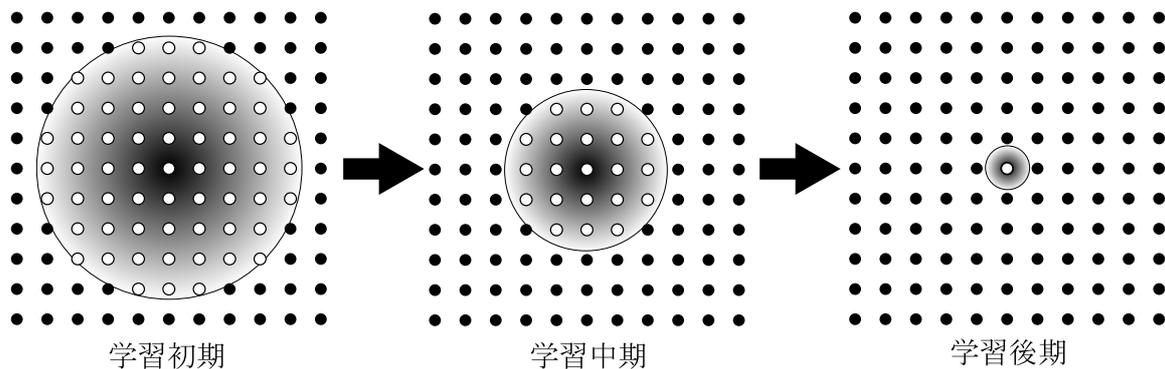


図 3.2 近傍関数で学習されるノードの遷移

ステップ 5. 繰り返し

学習回数 t が学習繰り返し回数 λ に達していなければステップ 2 に戻り、学習を繰り返す。

ステップ 6. 学習結果のマッピング

学習が終わった出力層に対し、入力データのマッピングを行う。入力層の各ノードの重みが、出力層のどのノードの重みに最も近いかを勝者ノードの選択と同じ方法で見つけだし、それぞれの勝者ノードの位置にマッピングする。このとき、近い位置にマッピングが行われた入力データ同士が似通っていると判断される。

3.2 球面 SOM

球面 SOM とは、SOM の出力層となるノードが球面上に配置されたもので、H.Ritter によって提案された [7]。ノードを平面に配置した場合との違いは、マップに端が存在しない

3.3 測地線ドーム

ことである。2.3.2 節で述べたように、マップに端が存在すると学習結果が歪む。その歪みを抑制するための方法として、データ構造としてマップの端にあるノード同士をつなげる方法があるが、この方法で実装したとしても、表示方法が平面だとユーザが端にあるノード同士の関係を読み取るのは困難である。

そこで本研究では、表示方法も球面にできる球面 SOM を用いてシラバスを可視化する。

3.3 測地線ドーム

球面 SOM を実現するためには、SOM の出力層となるノードを球面上に配置しなければならない。その際、ノード間の距離は科目間の位置関係が直観的に理解しやすいように等間隔にする必要がある。そこで、頂点数が多く 1 辺の長さが等しい多面体で球を近似し、その頂点にノードを配置する。

球に近似した多面体を構築する手法として、測地線ドームがある。測地線ドームは、球に内接する正多面体を初期状態として各辺を分割していくことによって構築される。具体的な構築の手順を以下に示す。

1. 初期状態となる球に内接する正多面体を用意する。
2. 多面体の各辺の中点を取る。
3. 各辺の中点を球面に接するように押し出す。
4. 押し出された点を新しい頂点として多面体を構築する。
5. 新しい多面体に対してステップ 2. からの手順を任意の回数繰り返す。

図 3.3, 図 3.4, 図 3.5 に構築の手順を任意の回数繰り返した状態を示す。それぞれのひとつ前の状態の多面体にある 3 角形の中に、さらに小さい 3 角形ができていることがわかる。

本システムで使用した測地線ドームは、分割を 3 回繰り返した状態のもので、頂点数は 642 個となっている。

測地線ドームは、多面体の中点を球面に押し出す過程で歪みが生じるため、すべての辺は厳密には等しくならない。この歪みを抑制するためには、より面数の多い正多面体を初期状

3.4 キーワードの分類

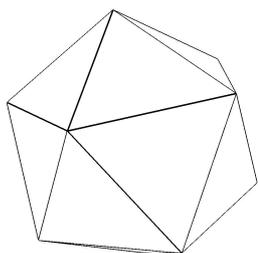


図 3.3 初期状態

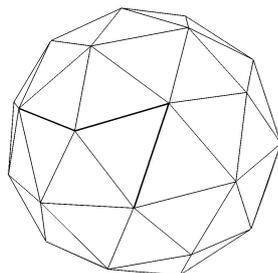


図 3.4 1回繰り返した状態

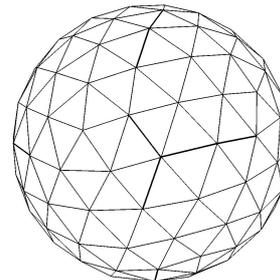


図 3.5 2回繰り返した状態

態として用いなければならない。そこで本システムでは、球に内接する正多面体の中で最も面数の多い、正 20 面体を初期状態として採用した。

3.4 キーワードの分類

シラバスからのキーワードの抽出には、先行研究で構築された情報抽出システム [1] を用いる。このシステムは、シラバス中に記述されている項目名の語彙リストを作成しておき、その語彙リストに記載されている項目名に対応した項目内容を解析してキーワードとその出現頻度を抽出している。語彙リストを表 3.1 に示す。表中で括弧で囲まれた文字は、その文字を含む場合と含まない場合の両方の表現をその項目の語彙として扱う。

先行研究の情報抽出システムでは、抽出の際にキーワードをソサエティという単位に分類している。ソサエティとは、「自ソサイエティの領域ならびに近傍領域における学問、技術の調査、研究、および知識の交換を行い、他ソサイエティと緊密な協力を保ちつつ、自ソサイエティの活性化を図り、学問、技術および関連事業の振興に寄与することを目的とする」もので、電子情報通信学会が定めた研究組織である [4]。ソサエティには現在、以下の 5 つが存在する。

- 基礎・境界ソサエティ

「電子」「情報」「通信」に関する基礎的な研究が行われている。

- 情報システムソサエティ

3.4 キーワードの分類

画像処理や情報処理についての研究が行われている。

- 通信ソサエティ

無線やネットワーク等通信に関する研究が行われている。

- エレクトロニクスソサエティ

電気回路などの電子・電気分野に関する研究が行われている。

- ヒューマンソサエティ

人に関するヒューマンインタフェースなどの研究が行われている。

このソサエティを基に科目中のキーワードを分類し、その割合を示すことでその科目で学習する分野をユーザが理解できるようにする。

そして、科目中のキーワードが属しているソサエティの割合を基に球面を色分けすること

表 3.1 シラバス項目の語彙

項目	語彙群
科目名	科目名, 授業名, 講義名, 科目名称, 授業名称, 講義名称, 授業科目名, 教育科目名, 開設科目名
単位数	単位数, 単位
開講時期	開講時期, 開講学年, 開講学期, 開講セメスター, 対象学年, 曜日・時限
担当教員	担当教員, 担当教官
教材	教材, 教科書・参考書, 教科書, 参考書, テキスト, 教材図書
授業目的	授業(の) 概要, 授業(の) 目標, 授業(の) 目的, 授業(の) テーマ, 授業のねらい, 講義(の) 概要, 講義(の) 目標, 講義(の) 目的, 講義(の) テーマ, 講義のねらい, 開講意義, 達成目標, 目的, テーマ, ねらい
授業内容	授業(の) 内容, 授業(の) 計画, 授業(の) 内容, 授業(の) 日程, 授業スケジュール, 講義(の) 内容, 講義(の) 計画, 講義(の) 内容, 講義(の) 日程, 講義スケジュール, 概要, 内容, 計画
評価方法	成績評価, 成績評価(の) 基準, 成績評価(の) 方法, 評価(の) 方法, 評価(の) 基準
履修上の注意	履修上の注意, 履修に当たっての留意点, 注意点, 注意
備考	備考, その他

3.5 科目情報の表示方法

で、同じ分野を学習する科目が集まっている領域を直観的に理解しやすいようにする。

3.5 科目情報の表示方法

ソサエティの割合を基に球面を色分けする場合、その色はソサエティに対応する色をそれぞれの割合だけ混ぜ合わせた色になる。例えば、赤に対応するソサエティと青に対応するソサエティが同程度含まれていれば、その周辺は紫に着色されることになる。赤か青どちらかの割合が強ければそれぞれに近い色合いになる。

球面上で類似している領域や科目間の関係を示すために、この方法は有効である。しかし、この色を見てユーザが科目にどのソサエティがどのくらい含まれているかを理解することは困難である。そこで、科目に含まれるキーワードが属しているソサエティの種類とその割合をユーザが直観的に理解できる機能を実装する。

先行研究 [6] において、キーワードの属性情報を帯グラフにして可視化するシステムが提案された。このシステムでは、キーワードの属性情報としてソサエティの他に、ソサエティよりもさらに細かい分類である、委員会コードや分野コードを用いて科目の情報を可視化していた。本システムで使用する属性情報はソサエティのみであるため、先行研究のシステムの内、科目のソサエティの割合を示す機能を利用して実装する。

3.6 可視化結果の表示方法

ユーザが可視化結果を見る際には、球体をある 1 方向から観測することになり配置されている科目の一部しか見ることができない。本システムでは、ある一部の科目間の関係ではなく、シラバス全体の科目間の関係を可視化しているため、可視化結果の全体像をユーザに把握してもらうことが重要である。そこで、ひとつの球体を表示する以外に球面の全体像を把握できる表示方法を実装する。

第 4 章

シラバス可視化システムの実装

4.1 システムの構成

本システムは、SOM のアルゴリズムを Java で、球面 SOM を表示するための機能を Java3D[5] を用いて構築した。

4.2 システムの概要

本システムは、ユーザからシラバスの文章を解析したファイルと、SOM の学習回数を入力してもらい、学習結果を球面に出力する。ここで、シラバスの文章を解析したファイルは、先行研究の情報抽出システムを使って作成されているものとする。実行画面を図 4.1 に示す。

図 4.1 の球体をマウスでドラッグすることで任意の科目が見えるようになる。また、球面上のノードが持つソサエティの割合を基に各ノードの色を決定し、ノード間の線分を始点に近づくにつれて始点のノードの色に、終点に近づくにつれて終点のノードの色に変化するようにした。

展開図を表示する機能などを実行した場合、球面上に表示される科目名が小さすぎて見えなくなる問題があったため、実行時の引数で科目名の文字を大きくできるようにした。

4.3 球面 SOM の実装

球面 SOM は 3.3 節で述べた測地線ドームを用いて実装した。そして、測地線ドームの頂点に配置したノードを入力層からランダムに選んだベクトルで初期化し、SOM のアルゴリ

4.4 単一科目の情報の表示

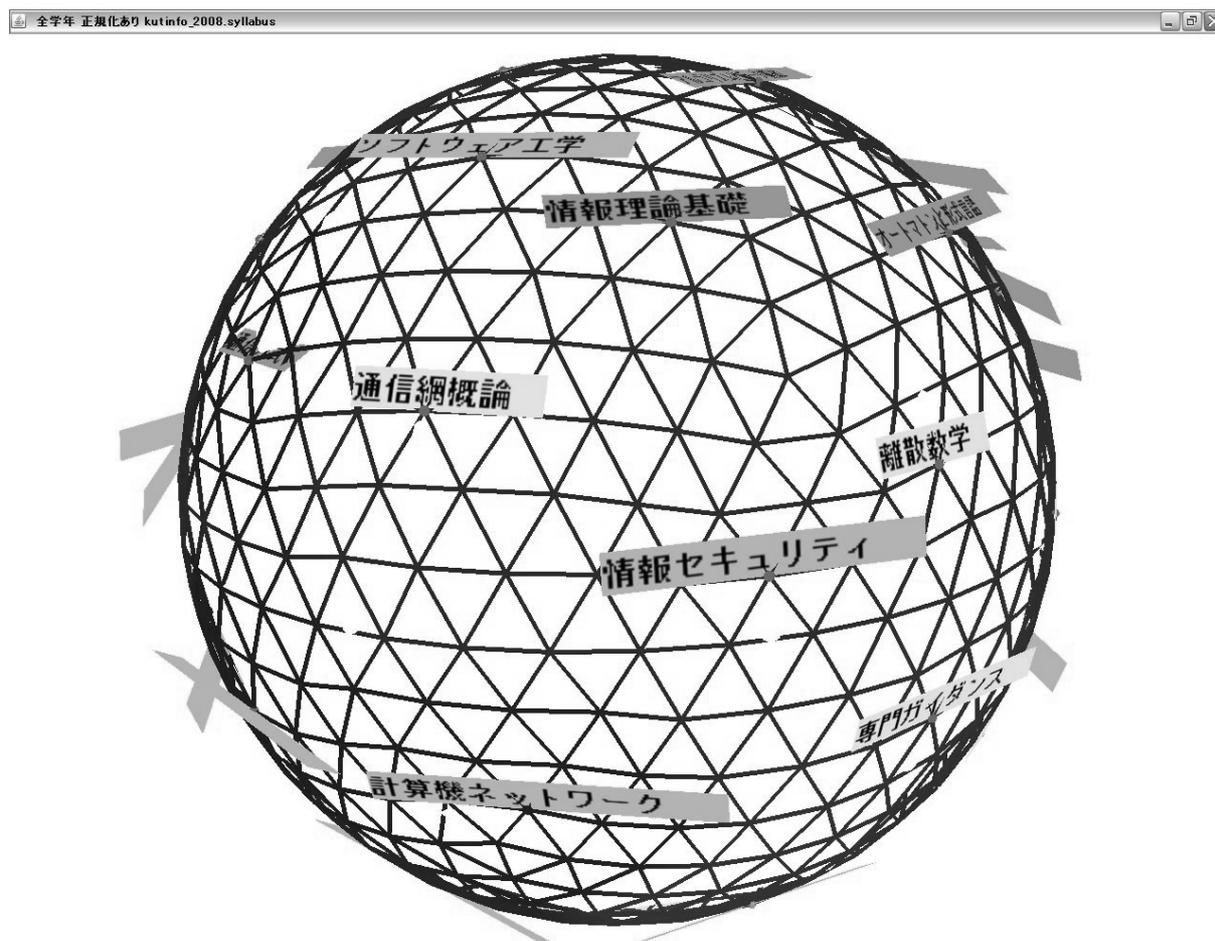


図 4.1 実行画面

ズムを適用する．ここで，入力ベクトルは先行研究と同じく，シラバスから抽出したキーワードの出現頻度を用いる．

4.4 単一科目の情報の表示

科目が持っているキーワードが属するソサエティの割合を帯グラフで表示する機能を，先行研究 [6] のシステムを用いて実装した．また，それぞれのソサエティに対応する色を表 4.1 に示す．

例として，「集積回路システム」に含まれるソサエティの割合を示した画面を図 4.2 に示す．

図 4.2 からは，「集積回路システム」にはエレクトロニクスソサエティに属するキーワード

4.4 単一科目の情報の表示

表 4.1 ソサエティと色の対応

ソサエティ	対応する色
基礎・境界ソサエティ	紫
情報システムソサエティ	赤
通信ソサエティ	青
エレクトロニクスソサエティ	緑
ヒューマンソサエティ	黄
どのソサエティにも属していない	黒

が最も多く、電子回路に関する分野を学習することがわかる。また、近くにある「論理回路理論」(図 4.3)よりもエレクトロニクスソサエティ以外のソサエティの割合が多いため、他分野への応用を学習する科目であることが読み取れる。

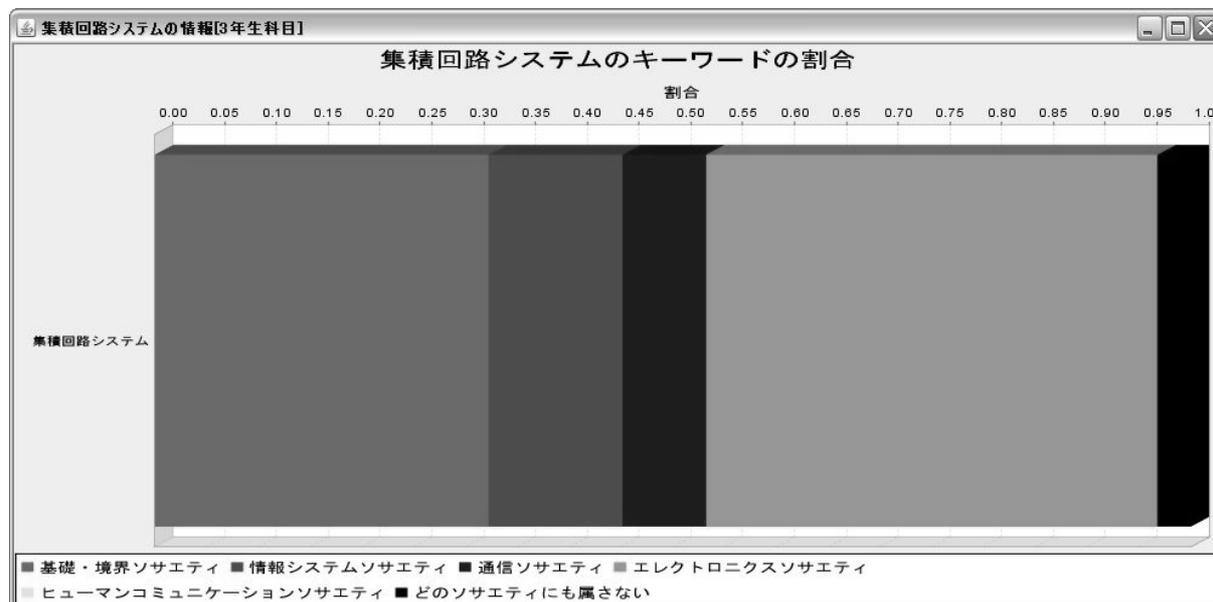


図 4.2 「集積回路システム」のソサエティの割合

4.5 ソサエティの割合によるベクトル間の距離計算

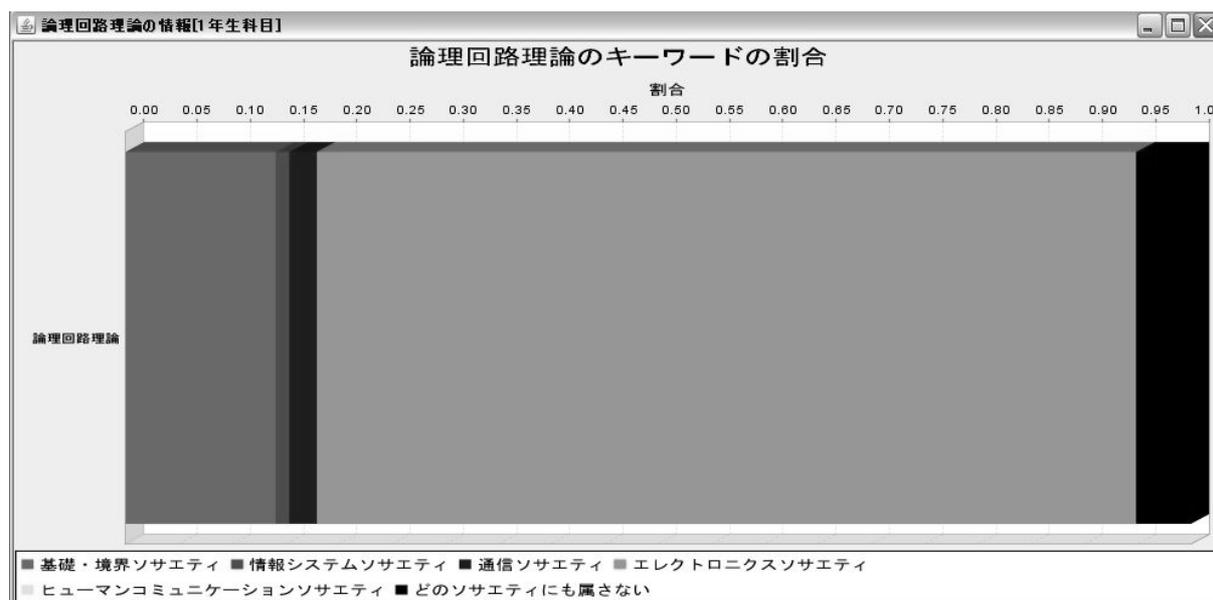


図 4.3 「論理回路理論」のソサエティの割合

4.5 ソサエティの割合によるベクトル間の距離計算

本システムでは、SOM のアルゴリズムのうち勝者ノードを決定する段階において、特別な処理をしている。本来、SOM では特徴を表すベクトルの距離だけで勝者ノードを決定するが、本システムではベクトルの各次元の関係性に着目した距離も用いる。

科目の特徴を表すベクトルは、出現したキーワードの数だけ次元数があり、各キーワードの出現頻度が各次元の値となっている。したがって、このベクトルの距離だけを計算すると、同一のキーワードを持っている科目だけが類似している科目と判断される。しかし、シラバスに含まれるキーワードはソサエティに分類されており、キーワードが属しているソサエティの情報は、類似している科目を判断するために重要である。例えば、「パターン理解」と「特徴抽出」はともに画像処理の分野で使われるキーワードであり、人間はこのキーワードが含まれている科目同士は似た分野を学習することを理解できる。しかし、別の言葉であるため SOM のアルゴリズムを適用すると類似しているとは判断されない。

そこで、本システムでは各科目にキーワードの出現頻度とは別にソサエティの割合を示すベクトルを持たせている。このベクトルは、各ソサエティに属しているキーワードの割合と、

4.6 科目検索

どのソサエティにも属していないキーワードの割合の6次元のベクトルである。そして、勝者ノードを決定する際に各科目が持つキーワードの出現頻度のベクトル同士の距離と、ソサエティの割合を示すベクトル同士の距離を足した値が最も小さいものを選んでいく。

4.6 科目検索

本システムでは科目が配置されるのが球面のため、ユーザに全体像が見えず、目的の科目がどこにあるのかが分かりにくい。そこで、科目名を指定するとその科目が視点の正面になるように球体を回転させる機能を実装した。

具体的には、科目名を書いたボタンが開講学年ごとに並べられたウィンドウを開いておき、そのボタンの中からユーザに検索したい科目名を選んでもらう。科目検索のウィンドウを図4.4に示す。

図4.5に科目検索を使って「論理回路理論」を検索した場合の画面を示す。

4.7 球の裏側と展開図の表示

球面の全体像を把握するための表示機能として、球の裏側からの視点を表示する機能と球の展開図を表示する機能を実装した。

球の裏側を表示する機能は、2つに区切った表示画面の左側に球の表側を、右側に球の裏側を表示させている。球の裏側の表示画面を図4.6に示す。

球の展開図を表示する機能を実装するために、まず3次元空間の座標を緯度と経度に変換する。XY平面上に球体の赤道があり、Z軸が北極に向かっているとき、3次元座標 (x, y, z) から経度 λ への変換は、

$$\lambda = \tan^{-1} \frac{y}{x}$$

となる。同じく、3次元座標から緯度 ϕ への変換は、

$$\phi = \tan^{-1} \sqrt{\frac{z}{x^2 + y^2}}$$

となる。次に、正距円筒図法を用いて緯度と経度を平面に投影する。正距円筒図法の投影式

4.7 球の裏側と展開図の表示



図 4.4 科目検索のウィンドウ

4.7 球の裏側と展開図の表示

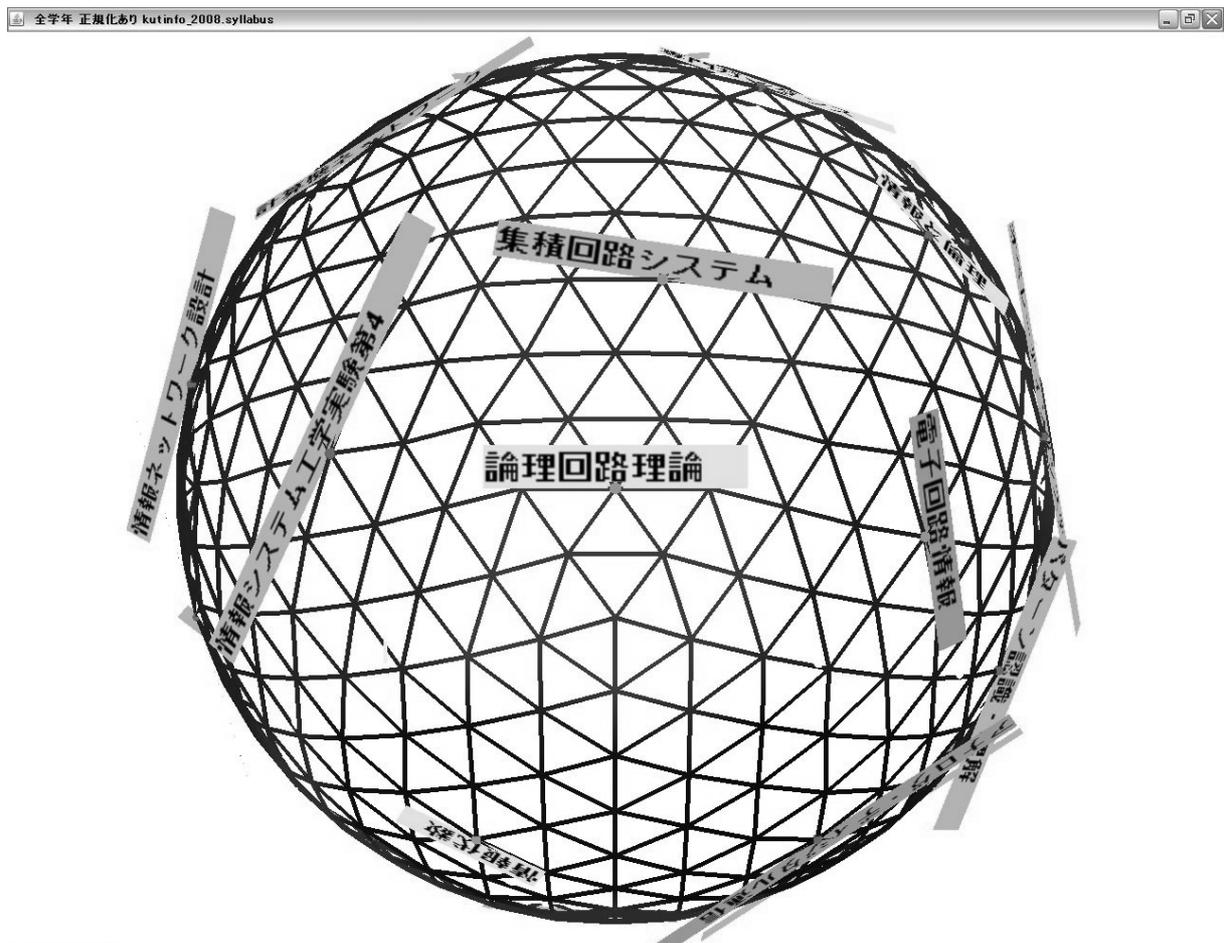


図 4.5 「論理回路理論」の検索結果

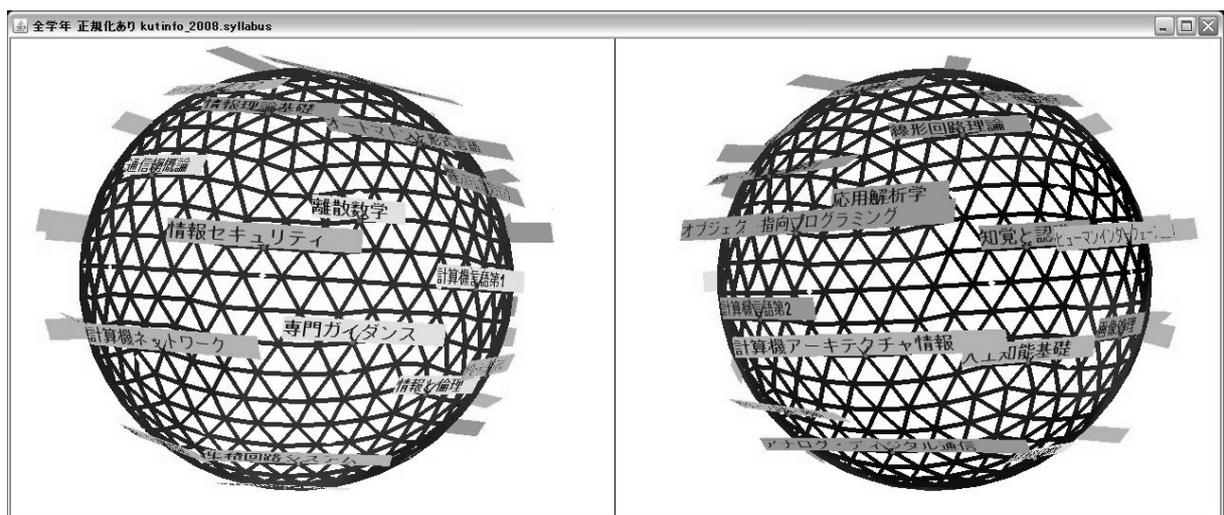


図 4.6 裏側の表示画面

4.7 球の裏側と展開図の表示

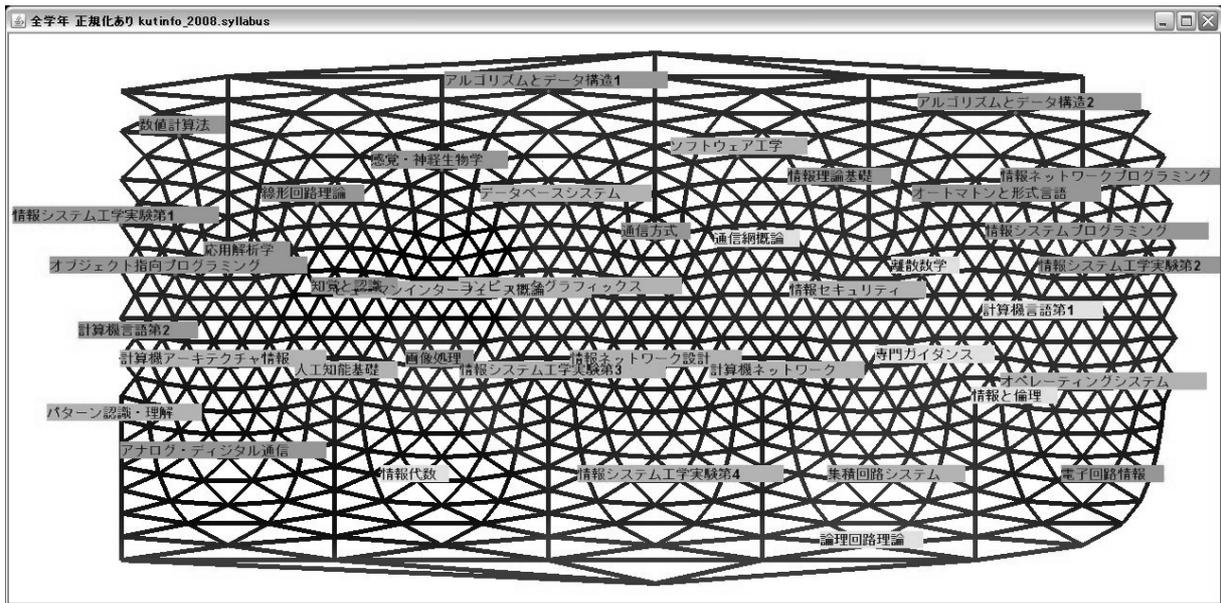


図 4.7 展開図の表示画面

は，緯度 ϕ ，経度 λ ，球体の半径を R とすると，

$$x = \cos \phi(R\lambda)$$

$$y = R\lambda$$

となるが，本システムでは球体の半径は 1，展開図の標準緯線は赤道を用いているので投影式は，

$$x = \lambda$$

$$y = \phi$$

となる．展開図の表示画面を図 4.7 に示す．

正距円筒図法の特徴として，展開図が 1 つの長方形になり，縦方向の距離が正確であることがある．海図として用いられるメルカトル図法も長方形になるが，各点の方向だけが正確であり，距離は正確ではない．さらに，メルカトル図法では極付近の点は表示することができないが，正距円筒図法なら極付近の点も表示することができる．

第 5 章

システムの評価

本章では，本システムの評価実験とその結果について述べる．さらに，評価結果から本システムについての考察を述べる．

5.1 評価方法

本システムに対する評価実験を，高知工科大学情報システム工学科の学部 3 年生以上の学生と，同じく情報システム工学科の教員を対象に実施した．評価方法は，まず被験者にシステムの概要と操作方法を説明した後，システムを操作してもらい，主観によるアンケート調査を行った．アンケート内容の性質から，被験者が可視化されている科目の内容をある程度理解している必要がある．そのため，すでにほとんどの科目を履修済みの学部 3 年生以上の学生と，シラバス作成者であり，実際にそれぞれの講義を担当している教員を被験者として選んだ．

評価アンケートの質問項目とその回答群を表 5.1 に示す．また，アンケート中で「科目に注目して」とある部分は，4.6 節で述べた科目検索の機能を使って科目が視点の正面になっている状態を表すこととする．また，設問番号の隣に示した図番号はその設問で被験者に見るように指示している画面である．実際には被験者は，球体を回転させるなどの操作をしながら画面を見ることになる．

5.1 評価方法

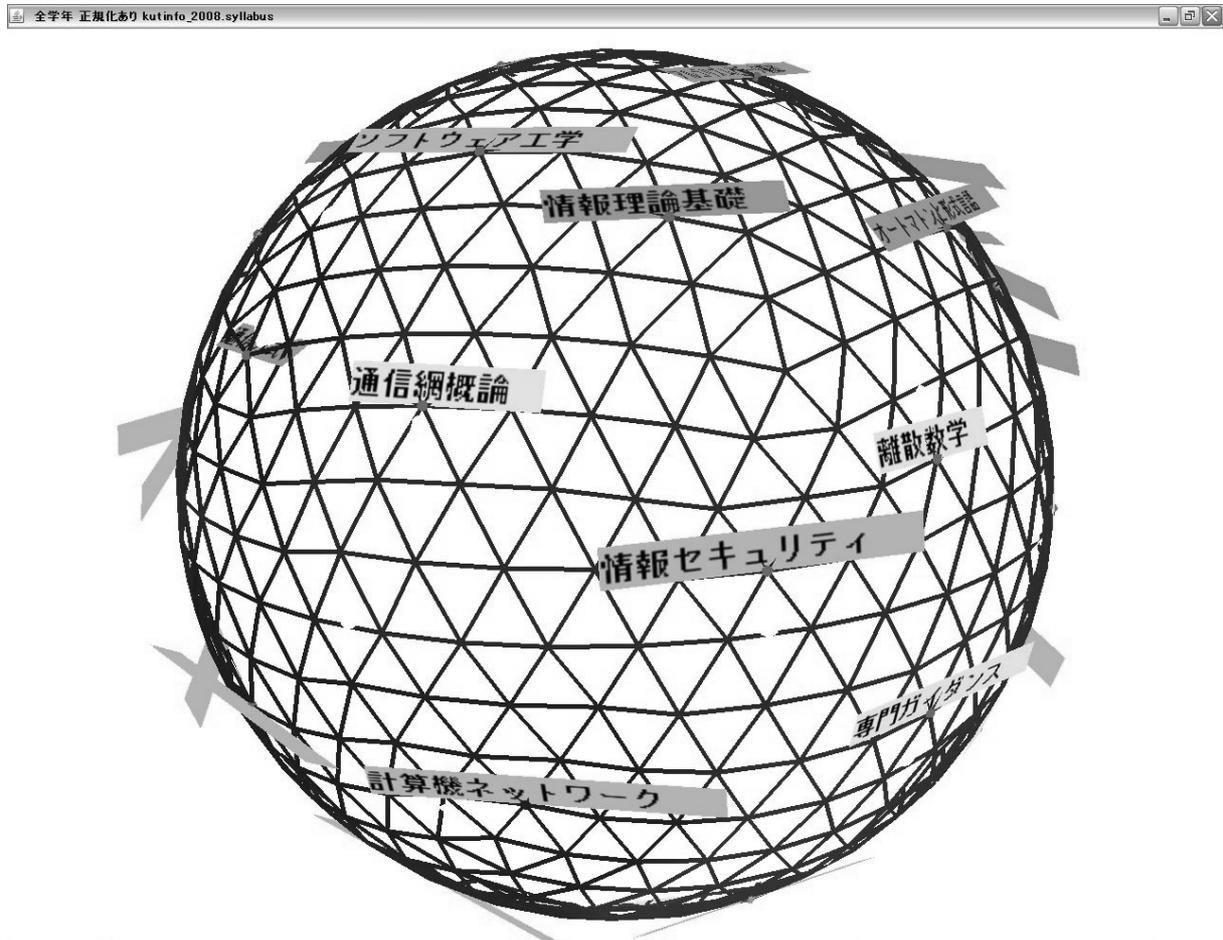


図 5.1 「問 1, 問 2」のウィンドウ

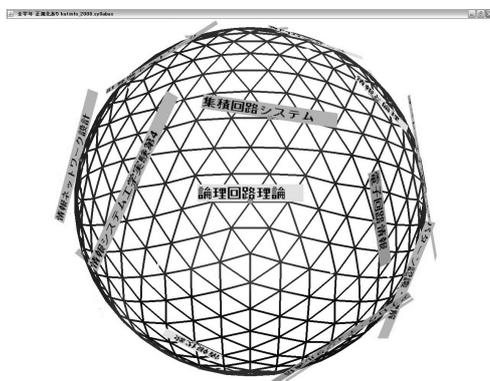


図 5.2 「論理回路理論」に注目した画面

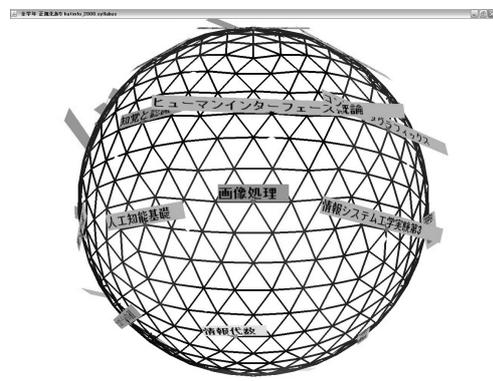


図 5.3 「画像処理」に注目した画面

5.1 評価方法

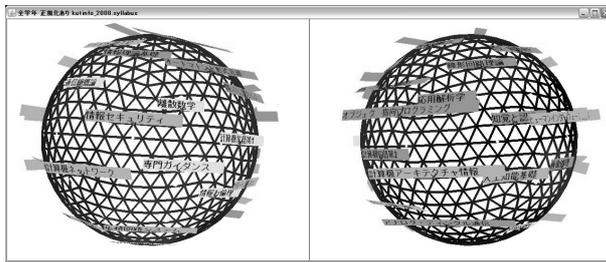


図 5.4 「問 3 . 裏側」のウィンドウ

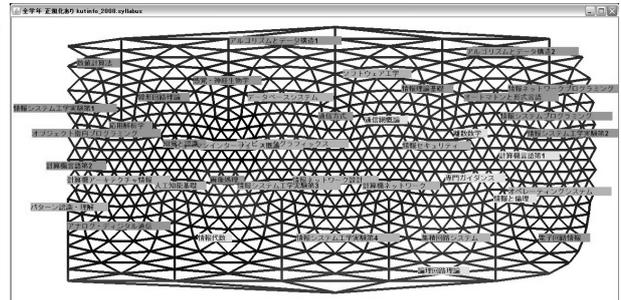


図 5.5 「問 3 . 展開図」のウィンドウ

Self-Organizing Map [10000/10000(終了)]					
応用解析学	情報理論基礎	計算機アーキテクチャ情報 オートマトンと形式言語		オブジェクト指向プログラミング 情報システム工学実験第1 計算機言語第2	ヒューマンインターフェース概論
パターン認識・理解	集積回路システム 電子回路情報 線形回路理論 専門ガイダンス	知覚と認識 情報代数		アルゴリズムとデータ構造2 アルゴリズムとデータ構造1	オペレーティングシステム
アナログ・デジタル通信		画像処理 データベースシステム 離散数学		情報ネットワークプログラミング 情報システムプログラミング 情報システム工学実験第2	通信方式 通信網概論
情報セキュリティ		人工知能基礎	ソフトウェア工学	計算機ネットワーク	
論理回路理論		コンピュータグラフィックス	数値計算法 感覚・神経生物学	計算機言語第1	情報と倫理
				情報ネットワーク設計	情報システム工学実験第4 情報システム工学実験第3

図 5.6 「問 4」のウィンドウ

5.2 評価結果

表 5.1 評価アンケートの質問項目とその回答群

設問	回答群
問 1 (図 5.1) (「問 1, 問 2」のウィンドウを見てお答えください) 任意の科目に注目したとき, 近くにある科目ほど 中心の科目と関係が深くなっていると思いますか .	1. なっていない 2. どちらかといえばなっていない 3. どちらともいえない 4. どちらかといえばなっている 5. なっている
問 2 (図 5.2, 図 5.3) (「問 1, 問 2」のウィンドウで「論理回路理論」と 「画像処理」に注目してお答えください) これらの周囲にある科目は中心の科目で 学習する内容を含んでいると思いますか .	1. 含まれていない 2. どちらかといえば含まれていない 3. どちらともいえない 4. どちらかといえば含まれている 5. 含まれている
問 3 (図 5.4, 図 5.5) (「問 3. 裏側」と「問 3. 展開図」のウィンドウを 見てお答えください.) どちらが科目間の位置関係がわかりやすいですか .	1. 裏側 2. 展開図
問 4 (図 5.1, 図 5.6) (「問 1, 問 2」と「問 4」のウィンドウを 見てお答えください.) どちらのシステムが科目間の関係を読み取りやすいですか .	1. 「問 1, 問 2」のシステム 2. 「問 4」のシステム

5.2 評価結果

評価実験のアンケートに, 学生 7 名, 教員 5 名の計 12 名から回答を得た. 評価実験の結果を表 5.2, 表 5.3, 表 5.4 に示す. 回答人数の内 () 内の数字が教員の回答人数である.

表 5.2 問 1, 問 2 に対する回答

	1	2	3	4	5
問 1 に対する回答人数	0(0)	0(0)	0(0)	10(4)	2(1)
問 2 に対する回答人数	0(0)	0(0)	2(1)	4(2)	6(2)

評価実験の結果, 「画像処理」や「論理回路理論」などその分野での基礎的な内容を学習す

5.2 評価結果

表 5.3 問 3 に対する回答

	1(図 5.4)	2(図 5.5)
回答人数	6(3)	6(2)

表 5.4 問 4 に対する回答

	1(図 5.1 のシステム)	2(図 5.6 のシステム)
回答人数	8(3)	4(2)

る科目を中心に見た場合，その発展科目との関係がよく表わされているということがわかった．以下に自由記入欄に書かれたコメントを抜粋する．

- 「ネットワーク設計」と「情報システム工学実験 3」が近いのでよい．
- 「画像処理」や「コンピュータグラフィックス」が近いのでよい．
- 「専門ガイダンス」や「情報と倫理」など，他と関連がなさそうな科目でも，近くに科目があることに違和感があった．
- 球面の裏側を表示すると，一番遠い科目がわかるのでよい．
- 2 科目以上の関係を見るときは展開図があるほうが好ましい．
- 操作が直観的でない．
- 色分けがわかりづらい．
- 図 5.6 のシステムでは科目間の細かい距離はわからないが，まとまりがよくわかる．

球体の裏側からの視点を見せる表示方法と球体の展開図を見せる表示方法では，表 5.3 で意見が分かれていることから，ユーザの見たい情報によって見やすいと感じる表示方法が異なることがわかった．操作性に関しては球を回転したときに科目名の向きがそろっていないと見づらいという意見が出た．

5.3 評価結果の考察

評価アンケートの問1の結果から、本システムで科目間の関係はおおむね良く表現できているということがわかった。しかし、コメントとしてある科目から一番近い位置にある科目でも直観的に遠いと感じるという意見があった。これは、システム上では近いと判断される距離が、人間にとっては遠いものであり、類似している科目として直観的に理解できないということである。あるノードとその最近傍にあるノードとの距離を1とすると、評価に用いた高知工科大学情報システム工学科のシラバスの可視化結果では、最も近い科目間の距離は3であり、被験者によってはこの距離を遠いと感じることがあった。

また科目によっては、関係ない科目が近くにあるように見えるという意見もあった。これは問2の設問であった意見で、「論理回路理論」と「情報システム工学実験第3」、「画像処理」と「情報システム工学実験第4」が近くにあることに対するものである。原因として、類似している分野の範囲が明確になっていないことがある。球体を1方向から観測すると、観測者からは円形の範囲が見えることになる。しかし、類似した科目が常に円形の領域に集まるとは限らない。特に「論理回路理論」は関係が強い科目が「電子回路」と「集積回路システム」の二つであり、三角形の領域を作りやすい。そのため、観測できる範囲に本来違う分野を学習するはずの実験科目が入り込んでしまい、ノイズとなっている。

この問題を解決するための手法として、本システムでは測地線ドームの各辺を色分けして類似している分野を示していたが、さらに分野の境界を明確にすることがある。分野の境界を明確にすることで、システムで類似していると判断された科目がユーザに理解しやすくなり、より正確な情報を提供することができる。また、前述したシステムにとっての近さと人間が感じる近さのずれを抑制するためにも、分野の境界を明確にすることは有効である。

問3の結果については、ユーザがどのような情報をシステムに期待しているかによって結果が分かれた。裏側を支持した被験者からは、ひとつひとつの科目を見比べる場合には適しているという意見があった。また、球面の裏側が見えることで、一番関係が弱い科目が見えることも特徴的であるという意見もあった。展開図については、球面上のすべての点の距離

5.3 評価結果の考察

を保持した展開方法がないため、全体像を把握しやすいメルカトル図法を用いて実装した。そのため科目間の距離が歪み、それを理由に裏側の方を選ぶ被験者もあった。

科目間の距離を正確に表現して球面を展開する方法として、正距方位図法がある。しかしこの図法では、展開図の中心となる点からの距離が正しいだけで、中心にない2点間の距離は歪んでしまう。結局、正距方位図法を用いた展開図で得られる情報は、球体を直接観測した場合と変わらないのである。正距方位図法を用いて展開図を実装しようとするならば、マウスなどの入力によって中心となる科目を任意に替えられる機能を追加しなければ実用的とは言えない。

問4の設問は、先行研究のシラバス可視化システムと本システムの比較である。結果として、本システムの方が科目間の関係性を読み取りやすいという評価を得た。これは、先行研究では類似した科目はひとつの矩形範囲にまとめられていたが、本システムではひとつひとつの科目が独立しており、より細かい科目間の関係を示していることが評価された結果であると考えられる。

しかし、本システムの課題として操作性の向上がある。先行研究より詳細な情報を表示できるようになったことで、ユーザが得たい情報を探す手間が増えてしまった。そこで、今後の課題として、ユーザが必要としている情報に限定して表示する機能の充実が考えられる。本システムに実装している科目検索の機能は、すでに可視化された状態の中から任意の科目がある場所を示すだけであるが、システムで類似していると判断された科目だけを新規に可視化するという機能があればユーザに煩雑な操作を強いることがなくなると考えられる。

また、本システムで対象としているシラバスは、Web上で公開されているものであるため、システムをWeb上で動作するアプリケーションとして実装することで、より実用性が向上すると考えられる。

第 6 章

まとめ

本研究では、先行研究において SOM のマップを平面で作成することによって生じていた歪みを抑制するために、球面 SOM を用いたシラバス可視化システムを構築し、その評価を行った。

可視化結果の表示には球体を多面体で近似した測地線ドームを用いることで、球面に SOM の出力層を実現した。可視化に用いるアルゴリズムでは、より正確な可視化結果を得るために、SOM の勝者ノードを決める際に科目が持つキーワードのソサエティの割合を加味するように実装した。また、ソサエティの割合によって球面を色分けすることで、類似した科目が集まっている領域を直観的に理解できるようにした。

評価実験では、教員と学部 3 年生以上を対象に、本システムでシラバスに記載されているが読み取りづらかった情報が正しく可視化されているか、先行研究と本システムの優位性についてアンケート評価を行った。評価の結果、本システムではシラバスの情報を正しく、先行研究よりも直観的に可視化できていることがわかった。

今後の課題として、類似領域の境界を明確にするなどの表示方法の改善や、より直感的に操作できるようにインタフェースの改良、情報の検索機能の充実が考えられる。また、より高精度の可視化結果を得るために、SOM 以外の手法の検討も必要であると考えられる。

謝辞

本研究のすべてにおいて、多大なるご指導を賜りました妻鳥貴彦先生に心より感謝いたします。

また、ご多忙の中、副査をお引き受けくださいました、岩田誠先生、吉田真一先生、本研究に対して貴重なご助言をくださいまして、ありがとうございました。

修士2年生の寒川剛志さん、大黒隆弘さん、ご自分の研究もお忙しい中、時間を惜しまずご指導にあたってくださいましたこと深く感謝いたします。

修士1年生の畠山博和さん、福田将行さん、藤原健太郎さん、山崎雄大さん、就職活動でお忙しい中、プレゼンの練習にお付き合いくださいまして、ありがとうございます。

一緒に卒研を最後までやり抜いた学部4年生の清水雅也さん、森拓也さん、竹内雄人さん、別府瞳さん、池田真実さん、西川貴仁さん、本当にお疲れ様でした。時には知恵を出し合い、励まし合える仲間に出会えたことを嬉しく思います。

システムの評価にご協力いただきました学部3年生の中澤大樹さん、濱野純平さん、細川恭平さん、前田晃宏さん、松井勇貴さん、松本直樹さん、ありがとうございました。特に濱野さんと松本さんは、3年生ながら精力的に活動されているのを見て、身が引き締まりました。

最後に、今までシラバス可視化に関わってきた先輩方、特にシステムの構築でも論文においても大変参考になる資料を残してくださいました木下聡さんに心から感謝いたします。

参考文献

- [1] 木下 聡, “自己組織化手法を用いたシラバスの可視化に関する研究”, 高知工科大学 平成 19 年度修士学位論文, 2007 .
- [2] Teuvo Kohonen, 徳高 平蔵, “自己組織化マップ”, シュプリンガー・ジャパン株式会社, 2005 .
- [3] 中塚 大輔 等, “球面 SOM の性質とその応用例”, 信学技法, NC2004-23, pp.67-72 , 2004 .
- [4] “電子情報通信学会”, <http://www.ieice.org/jpn/index.html>.
- [5] “Java3D”, <http://java.sun.com/javase/technologies/desktop/java3d/index.jsp>.
- [6] 加集 広希, “キーワードの属性情報を利用したシラバス情報の可視化手法の検討”, 高知工科大学平成 19 年度プロジェクト研究報告書, 2007.
- [7] H.Ritter, Self-Organizing Maps on non-euclidean Spaces, in Kohonen Maps, ed.E.Oja and S.Kaski, pp.95-110, Elsever, New York, 1999.