

感情表現を目的とした 教育用文章作成支援システムの提案

本田 隼生

指導教員 白濱成希

教育の分野にコンピュータを使う機会が増えている。しかし単語に含まれている感情を支援するシステムは少ない。よって本研究ではあらかじめ準備した単語に喜び・悲しみ、怒り・恐れ、期待・驚き、受容・嫌悪の4つの対になる感情ベクトルを割り当て自己組織化マップで表現することで、必要な感情に対する単語を提供する支援システムの構築を目的として進めた。作文に関する教育支援を想定し、修学旅行というテーマを定め、関連する単語を設定し、その感情の値はアンケートによって決定した。作成したシステムを被験者に実際に使ってもらい、入力した感情の値と近い感情をもつ単語が現れたか、マップ上での単語の位置が正しいかアンケートを行った。アンケートや課題、単語の感情値を決める際に現れた問題点に対して考察を行った。

1. はじめに

コンピュータの発展は目覚ましいものがあり、様々な分野や状況で活躍している。教育の分野においても e-learning などの普及によって学習支援にコンピュータを使用する機会の増加は著しい。それに伴い教育現場におけるコンピュータに対する要求も増大している。ひとつの単語を入力すると類義語や対義語を表示するものなどがあり、例を挙げてみると、翻訳サイトや電子辞書などがこれにあたる。また、単語に対してその意味を表示し支援するシステムは比較的多く存在するが、その感情的な情報を提示し支援を行うものはまだあまり普及していない。通常の情報処理の分野では感情や感性を取り扱うことが難しい。一般に言葉から受ける感情は人や状況などによって変化することが多く、曖昧性が存在するから困難であるためと考えられる。

主観性や曖昧性は一般に計算機で取り扱うことが困難である。このような場面でよく用いられるものにニューラルネットワークやファジィ推論などがある。曖昧な表現をコンピュータ上でも可能にすることを目的として作られた表現方法なので、コンピュータへの感情導入の分野でも十分実用可能といえる。その中でもニューラルネットワークの一つである、SOM は可視性についても優れていることや多次元のものでも処理することができる。本システムの構築にあたり求める感情に対し候補となる単語群を表示する必要がある。勝者ベクトルが近傍にあたり、それを可視化することができる SOM は、システム構築に適していると思われるので採用した。

SOM を採用することで、主観性や曖昧性などの問題をある程度解決することを目的とする。またシステムが妥当かどうかを判断するために、小学校高学年を対象とした、感想文を作成する最中に行き詰っ

た場合、現段階での感想文に必要な感情の数値を入力すると、感性情報から被験者に適していると思われる単語を支援するというを想定したシステムを構築し実験を行う。

2. 感情のモデリング

本研究では感情を工学的に取り扱うことを可能とするために感情のベクトル空間での表現を行う。感情空間の次元数は感情心理学における様々な説に基づいている場合が多い。よって Pultchik の理論を採用し、ベクトルによる表現を行った。

2.1 Pultchik の感情理論

Pultchik は「日常生活の中にみられる複雑な情動は、いくつかの因子に分けられ、これはまた総合もできる」と提唱した。

Pultchik は多次元尺度構成法によって 8 つの基本感情とその属する感情を立体化した。この立体モデルは横の断面には 8 つの基本感情が配置され、縦の次元では強度を表している。また次のような 6 つの公準によって作られている。

- (1) 数個の基本(一次)感情がある。
- (2) 基本感情の結合の仕方によって複雑な混合感情が生まれる。
- (3) 各基本感情は、その生理的行動において特異である。
- (4) 日常生活で普通われわれのしている感情は混合感情であり、基本感情は混合感情から推測される。
- (5) 基本感情は両極端におかれ、2 つの感情がそれぞれ対になって位置づけられる。
- (6) 感情はそれぞれ強度差あるいは覚醒差がある。

本研究で用いる 8 つの基本感情のモデルを図 1 に示す。

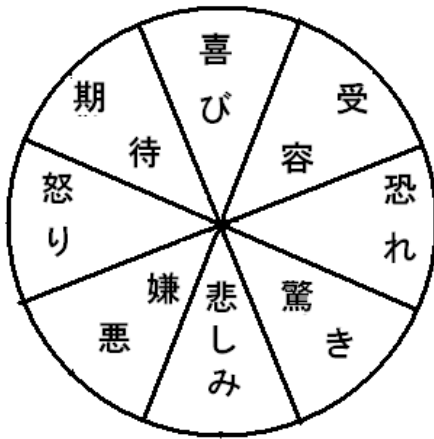


図 1: Plutchik の基本感情モデル

2.2 単語のベクトル表現

公準(1)の基本感情を以下の直交規定ベクトルに対応付ける。

e_1 : 喜び-悲しみ e_2 : 怒り-恐れ

e_3 : 期待-驚き e_4 : 受容-嫌悪

今回、作文のテーマとなる単語を感情ベクトルで表現する必要がある。単語のベクトルを \vec{W} とすると

$$\vec{W} = W_1 \vec{e}_1 + W_2 \vec{e}_2 + W_3 \vec{e}_3 + W_4 \vec{e}_4$$

で表わされる。係数 $W_1 \sim W_4$ をどのように決定するかについては、様々な方法があるが、今回は十分な数の被験者を確保し、11 段階評価で割り当てた値を使用した。同研究室内の他の研究では正規化された感情のデータベースを使用しているが、本研究ではそれとは別に独自に集めた単語において数値を割り当てているので矛盾はしない。ほかの研究とリンクするときは正規化を行うことで問題は解決する。

3. SOM について

では、今回の実験に使用した SOM について説明する。

3.1 SOM の概要

SOM (Self-Organizing Maps) とは、Kohonen によって開発された人間の脳の情報処理方式を工学的に表した数理科学モデルの一つで、自己組織化マップのことを指す。入力パターン群をその類似度に応じて分類する能力を自律的に獲得していく、教師なし学習の階層型ニューラルネットワークである。入力データがどの分類カテゴリーに属するかは与える必要がなく、元の空間における順序性や類似性などの相互関係をニューロンの二次元的なつながりを持つマップにすることができる。

その特徴から様々な分野で SOM は利用されている。例えば、科学分野におけるの定常分析や定

量分析、大量のデータからの特徴抽出などである。他にも、多次元のデータを二次元にできることから、巡回セールスマン問題などに対する最適化手法としても利用されることがある。

また SOM には、改善の加えられた種類が多く存在する。代表的なものを挙げれば、SOM が教師なし学習であるのに対して、教師あり学習にした LVQ や学習順序に依存する性質を改善されたバッチ学習 SOM、非ベクトルなデータへの応用が可能な中央値 SOM、球面状にマッピングを行うことで学習の偏りを軽減した球面 SOM などがある。

3.2 SOM のアルゴリズム

先ほど述べたように SOM は階層型ニューラルネットワークであり、図 2 に示すように二層のネットワークを持つ。

第一層は n 次元の入力層 $x(t)$ であり、第二層は競合層と呼ばれ、出力を視覚的にするため二次元配列が一般となっている。競合層のベクトルは、参照ベクトル $m_i(t)$ で表現され、入力層の n 次元と同様に n 個の要素を持つ。学習のステップはまずすべての参照ベクトル m_i の要素をランダムに決定し、入力ベクトル $x(t)$ を与える。このとき、 $x(t)$ とのユークリッド距離 $|x - m_i|$ を最小にするようなニューロンを c とすると、参照ベクトル m_c を持つニューロンを勝者ユニットとする。勝者ユニット、およびその周辺の近傍 N_c 内のユニットは次式に従って入力ベクトルを学習する。

$$m_i(t+1) = m_i(t) + h_{ci}(t) [x(t) - m_i(t)]$$

また、 t は離散時間座標であり、 $h_{ci}(t)$ は学習率係数を含めた近傍関数のことを指す。

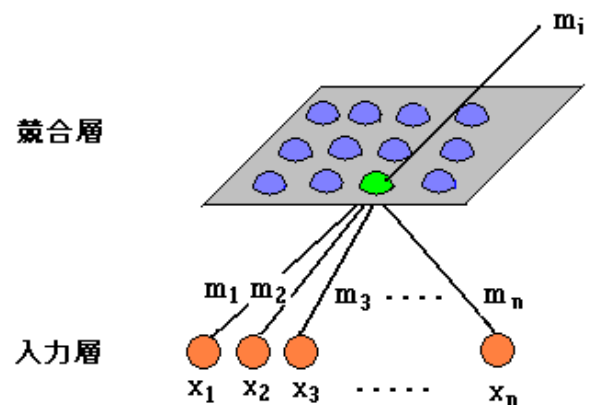


図 2: SOM の構造概念

まず入力信号が提示されると、その信号に一番近いユニットが判別される。それが m_c (勝者ユニット) として決められる。そして勝者の周りには近傍領域が図 3 のように正方形の中に囲った正方形のように定義される。SOM のマップに使われる学習近傍の形状には 6 角格子型と正方形型があり、一般的

に 6 角格子型のほうが見やすいとされているが、この部分では正方形近傍を用いる。

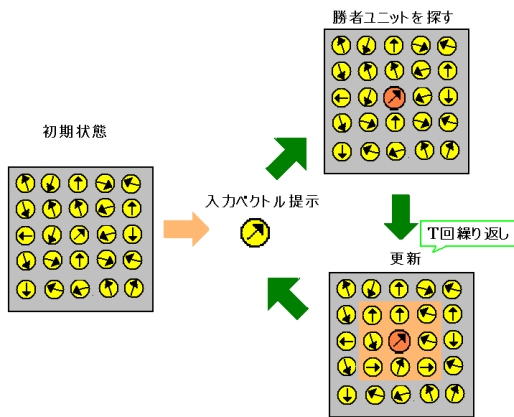


図 3: SOM の学習のステップ

勝者ユニットの周りに近傍が定義され、近傍内のすべてのユニットは上式に従って図のように近傍内の全てのユニットは入力ベクトルの方向に少し動く。

これを T 回繰り返して学習を行う。近傍サイズは $N_c = N_c(t)$ という時間の関数で表され、学習とともにそのサイズを小さくしていく。学習率係数の値においても更新回数が進むにつれて減少していく。すべての入力ベクトルに対して、これを繰り返すことにより、各入力ベクトルに類似したユニットが集まる。

4. システム概要

本研究のシステムは以下の二つのプロセスから構築されている。

4.1 SOM の作成による単語の分類

単語を感情の値によって分類するので、一つの単語につき 4 つのベクトルを用意した。それらを単語が持つ感性情報を表すとし、Pulchik の感情理論にならってそれぞれを喜び-悲しみベクトル、怒り-恐れベクトル、期待-驚きベクトル、受容-嫌悪ベクトルとする。以上のベクトルを今回の研究で用いる感情ベクトルとし、これらを基本に各ベクトルの組み合わせや度合いによって単語の分類を行う。今回は「修学旅行」をテーマにし、喜び、悲しみ、怒り、恐れ、期待、驚き、受容、嫌悪の感情を多く持つであろう単語を同研究室の学生に手伝ってもらい、それぞれ 10 個用意し、更に 57 個の修学旅行の感想文でよく用いられる単語や修学旅行と関わりのある単語を加え計 137 個の単語を用意し、Web クラス上でアンケートを用意し本校生に回答を行ってもらうようにした。テーマに「修学旅行」を選んだ理由は実験対象となる小学校高学年が感想文として書くこ

とが考えられることや、ほぼ全ての人が経験したことがあるためアンケートに回答しやすいと思ったからである。

アンケートを回答するに当たり、回答者に小学校高学年を想定して回答を行ってもらう。一つのベクトルにつき、評価を-5~+5 の 11 段階で割り当てもらう。実際にアンケートの際に用いた型を図 4 に示す。

喜び-悲しみベクトルなら+5 が喜びを強く表し、-5 は悲しみを強く表している。0 の場合はどちらも表していないことを示している。喜びと悲しみで一つのベクトルであるため、喜びがある場合は悲しみが無いものとする。ほかの三つのベクトルでも同じことがいえる。

	-5	-4	-3	-2	-1	0	1	2	3	4	5
悲しみ⇄喜び	○	○	○	○	○	○	○	○	○	○	○
恐れ⇄怒り	○	○	○	○	○	○	○	○	○	○	○
驚き⇄期待	○	○	○	○	○	○	○	○	○	○	○
嫌悪⇄受容	○	○	○	○	○	○	○	○	○	○	○

図 4: アンケート回答に用いた型

アンケートで回答してもらった感情のベクトル値を平均した数値を今回の SOM で用いる単語の感情ベクトルとする。得られた言葉の感情ベクトルを SOM によって二次元マップにし、単語の分類を行う。今回使用する SOM のプログラムは SOM_PAK-3.1 のものを用いている。

また、SOM によって多次元ベクトルを二次元マップにする際に SOM_PAK-3.1 に含まれている監視用プログラム「visual」を使用し、単語群の二次元マップ上での x 座標、y 座標を得る。

4.2 文章支援システムの構成

二つ目は全段階で得られた SOM を用いたシステムの構成である。必要な感情ベクトルの値を同様に-5~+5 で入力すると、その感情ベクトルの値に近いベクトル値を持つ単語が次式によって 5 つ選出される。単語が持つベクトルの係数をそれぞれ W_1, W_2, W_3, W_4 とし、入力された数値をそれぞれ a_1, a_2, a_3, a_4 、単語のベクトル値と入力ベクトル値の差を d とすると、入力されたベクトルと i 番目の単語の差を示す d_i は、

$$d_i = (W_{i1} - a_1)^2 + (W_{i2} - a_2)^2 + (W_{i3} - a_3)^2 + (W_{i4} - a_4)^2$$
 となる。入力された値と全ての単語を比較した後、ソートを行い最小の d を見つけ、その単語を表示する。同様の動作を 5 回繰り返し、入力された値に近い値を

持つ単語を 5 個表示する。本来は距離を考慮する必要はあるが、距離によって十分な検索結果が表示されない恐れがあるので距離を考慮しないものとする。

表示された言葉の中から 1 つを選択するとその言葉が、アンケートの結果から得られた SOM 上での位置にあるかを SOM 作成時に「visual」によって得られた座標データベースをもとに表示される。表示された位置から単語を探し近い単語を確認して、使用する単語を選ぶというシステムである。

5. アンケート結果

アンケートにより言葉に割り当てた感情のベクトルによって SOM を作成し分類を行った。その結果の全体像を図 5 に示す。

なお、SOM の二次元マップは、元の次元のつながりによって構築されるが、先ほど述べたように学習回数やマップの大きさなど様々な要因で結果が大きく左右されてしまう。そこで今回はマップの種類を次の様に設定し分類を行った。

- ユニット数 40×40
- 近傍関数 ステップ関数
- 近傍関数の形状 六角格子型
- 学習回数 5000
- 学習近傍領域の初期値 6
- 初期学習率係数 0.05
- 学習率係数の関数 線形関数

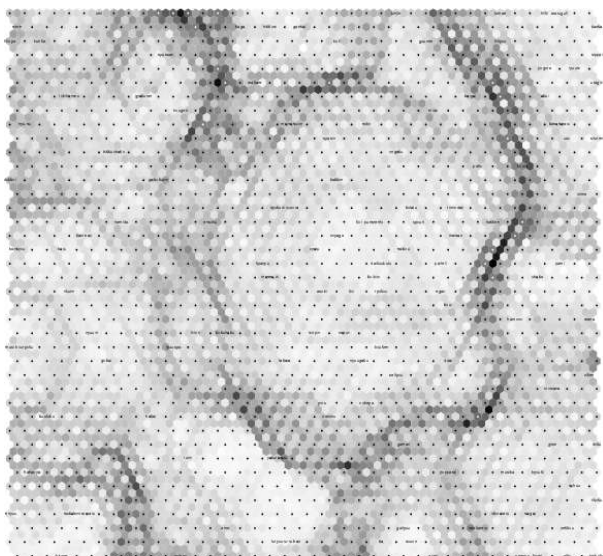


図 5: アンケートから得られた SOM

中心には喜びと期待の感情を多くもった言葉が集まっており、右上には怒りと嫌悪をもつ言葉、右下には、悲しみ、恐れ、嫌悪をもつ言葉、左上には

喜びと驚きをもつ言葉、左下には喜びと受容をもつ言葉がそれぞれ主に集まった形となった。

6. まとめ

図 4 を用いてシステムの作成を行った。ただし、このシステムでは小学校高学年を対象とすることを目的としているが、本研究で用いた SOM_PAK-3.1 の SOM では日本語表記ができないため、分かりにくい、もしくは理解されない恐れがある。よって図 5 に載っている単語を全て日本語に書き直したものをシステムでは使用した。実際のシステムを図 6 に示す。

右部分が図 5 の単語を日本語にして見やすくした SOM の二次元マップである。左上に位置しているのが入力部分である。そこに数値を入力、もしくはバーをスライドすることで入力値が変化する。その下にあるのが、今回の実験で使われる単語群とアンケートより割り当てられた数値である。ここにあるベクトル値や単語を変化させてもシステムに影響はされない。この部分は、どのようなシステムで使用されている単語や、単語の感情値の度合いを分かりやすくするために用意した。入力部分に数値を入れ表示ボタンを押すと、下の部分に入力したベクトル値と近いベクトル値を持つ単語が集まる。集められた単語を選択すると最下層に単語の座標が表示されるというシステムである。

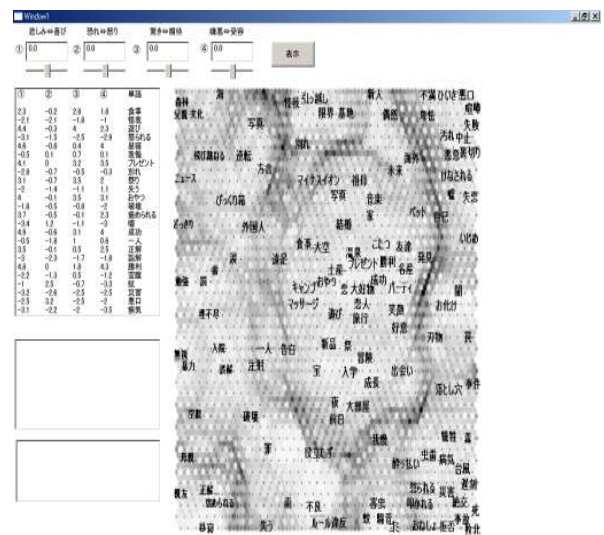


図 6: 文章作成支援システム

作成したシステムを検証するために実際に被験者に使ってもらい、「入力した値と近い単語が現れたか」、「SOM 上での単語の位置は正しいか」の二つについて本校の学生 30 名を対象にアンケートを取った。

一つ目のアンケート、「入力した値と近い単語

が現れたか」では「現れた」、「大体现れた」と回答してくれた方が約三分の二いたのでそれほど外れていない結果となったといえる。(図7)

しかし、信頼できる数値とまではいえない。精度をよりあげるためにはアンケートの人数を増やすことで、単語一つに対してより多くの人の感情ベクトルを入れる必要があると感じた。

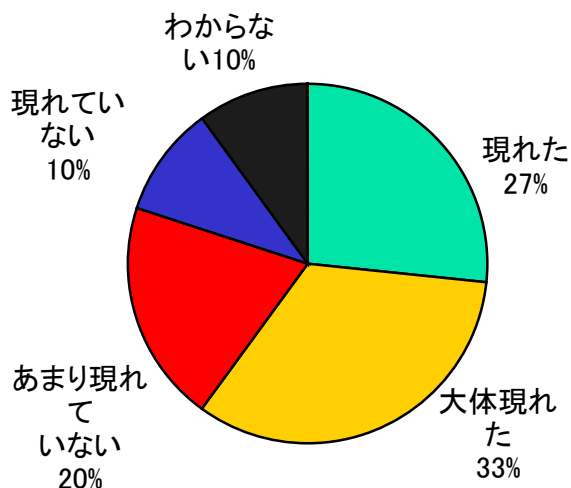


図7:入力した値と近い単語が現れたか

二つ目のアンケート、「SOM 上での位置が正しいか」は「正しい」、「大体正しい」と答えた方が約半数、「わからない」と答えた方が全体の約四分の一。(図8)

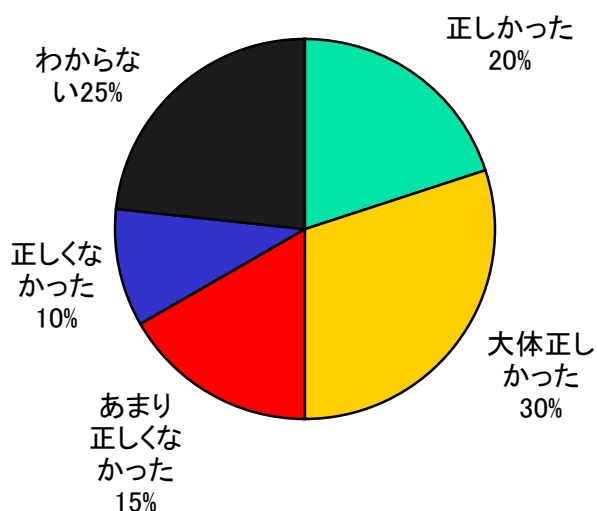


図8:SOM 上での単語の位置が正しいか

わからないと答えていた方は図5における中心部分や右下の部分の単語を選択していることが多く、それ以外の部分の単語を選択している人には少なかった。中心部分や右下部分はほかの場所と比べると単語の密集度が違うことがわかる。(図9)

単語と単語の間の色が薄ければ薄いほど、また距離が近くなるにつれ二つの単語は近い感情のベクトルを持つことをSOMは示している。しかし、距離が近くても間の色が濃かったり、距離が離れていたりするほど、二つの単語が持つ感情のベクトルは似ていないことを示している。

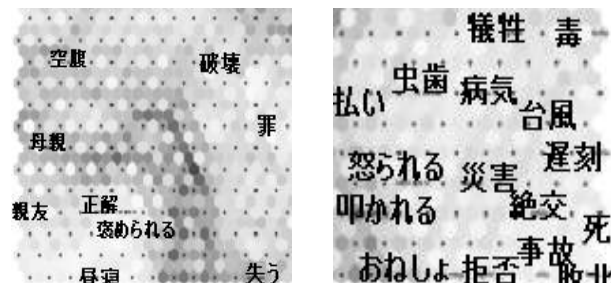


図9:SOMの通常部分(左)と密集部分(右)

しかし、図9の右図のように文字が密集している場所では、通常部分(図9左図)では近いといえる距離でも大量の単語があるため、どこまでが感情ベクトルが近い単語と認識してよいか分かりにくく、また間色の判断も難しくなる。そのため「わかりにくい」と答えた方が多かったのではないかと考えられる。

7. 実行例

本システムが必要とされる状況は感想文を執筆するにあたって筆が止まる場面などが考えられる。例えば「今日は修学旅行があった。」からと始まる文章があるとす。適切な言葉が思いつかない時や、自分とは違う感性の単語が必要となる時に、今回のシステムを用いて単語の支援を行ってもらう。

例の中では「特に楽しかったことは・・・であり、他にも」の次につながる単語を探すこととし、システムを実際に行う。入力部分に関しては改善の必要があるが今回は以下の感情ベクトルデータが与えられたものとする。

- ・ 喜び-悲しみ 3.6
- ・ 怒り-恐れ -0.5
- ・ 期待-驚き 3.3
- ・ 受容-嫌悪 0.3

与えられた感情ベクトルと近い感情ベクトルをもつ「前日」「新品」「キャンプ」「夜」「入学」が選出された。二次元マップ上での位置を調べるため、選出された単語の一つ、「前日」を選択すると、「前日」の座標は(21,29)であることが導き出された。今回のシステムでの二次元マップの最大座標は(40,40)であるため、「前日」は中心の下に位置していることが分かる。

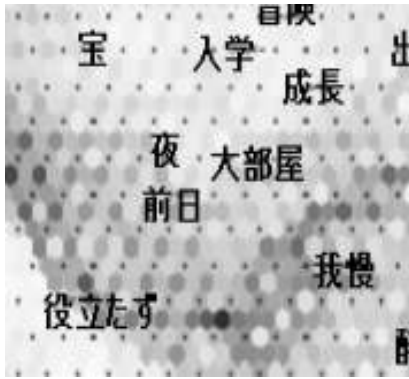


図 10:座標 (21,29) の拡大図

得られた二次元マップ上での座標から「前日」を探し出し、付近に位置している単語を確認すると、「夜」「大部屋」「入学」「成長」「我慢」「宝」「役立たず」があった。しかし、図 10 をみると「前日」と「我慢」や「役立たず」の間の色は濃くなっているため「前日」とその二つの単語の感情ベクトルは離れていることが分かる。そのため、それ以外の単語がシステムから得られた感想文を支援する単語となる。「他にも大部屋なので夜はみんなと語り合い…」 「入学当初と比べれば成長したと感じ、…」 など、上記の様な作文の支援が行えるといえる。

8. 今後の課題

今回の実験にあたって課題が見つけた。以下に課題とその課題に対する対処法を記す。

8.1 アンケート方法の見直し

後半に追加した単語はアンケートをしてもらう人数が前半に比べて少なかった。アンケートから得られた数値を平均化して単語の感情ベクトルとして扱うので、人数が少ないと偏りや別の単語と近似が起きてしまう恐れがある。また、前半に用意した単語は特にテーマが決めずに設定したため、後にテーマを定めると感情の値にばらつきがみられる単語があった。例として「夜」が挙げられる。テーマの設定がない場合にアンケートを行うと、「夜」では恐れや嫌悪の感情値が高かったが、修学旅行というテーマを定めると恐れや嫌悪の感情値は減少し期待や喜びの感情値が高丸という結果となった。今後の研究では、まずテーマを決めてからそれに関連する単語を用意し、随時単語を増やしていく方法を用いるべきである。

8.2 ベクトルに用いる感情の見直し

単語に感情のベクトルを四つ割り当て、喜びと悲しみ、怒りと恐れ、期待と驚き、受容と嫌悪をそれぞれ一つの対になるベクトルとして割り当てていたが、アンケートを受けてもらう際に、ベクトルに対応して

いる感情が相反していないのではないかという意見を多くもらった。今回の研究ではあらかじめ研究室で使われていた感情のベクトルをそのまま使用したため、感情の部分进行深入していなかった。個人にとっては相反していない二つの感情が一つのベクトルとなっていることも分かりにくさの原因と考えられる。感情の部分についてももう少し理解してから感情のベクトルを用意するべきであった。

8.3 単語収集方法の見直し

テーマを決め、単語を集めるだけでは似たような意味を持つ単語が集まり、近い感情ベクトルを持ち SOM 上での距離が近くなる。すると他の単語との区別がつかなくなり、システムとして意味を成さなくなる恐れがある。そこでテーマを決めた際に感情別に単語を集めることで意味や感じ方が近い単語の使用をなるべく控えるべきであると感じた。

8.4 システム入力部分の見直し

今回のシステムでは感情を-5~+5の数値で入力してもらう仕様にしてしたが、感情を数値で表現することは実生活ではほとんどみられない。よって感情の数値化に慣れている人は少なく、理解されにくいという恐れがあり得る。そこで対処法として、数値入力部分を、感情を表現する言葉が集められた SOM による二次元マップなどの図形にしマップ上の単語の選択を入力とすることで、より直感的な入力部分にする方法が考えられる。

8.5 支援方法の見直し

今回のシステムでは、入力されたベクトルと値の近い感情ベクトルを持つ単語を選択すると SOM 上での単語の位置を座標で表すようにしている。中心から外れた場所は見つけにくくなる可能性がある。そこで対処法として、選択された単語の座標を中心とし、マップを拡大した画像を表示させることが考えられる。

9. 謝辞

最後に本研究を行うにあたり、ご指導をいただいた白濱教員や、様々な意見を述べてくれた同研究室の方々、アンケートに回答していただいた本校生の方々に感謝の意を表したいと思う。

参考文献

[1] 徳高平蔵・岸田悟・藤村喜久朗、『自己組織化マップの応用 多次元情報の2次元可視化』, 海文堂(1999)

[2] T.Kohonen, 『自己組織化マップ 改訂版』, シュプリンガー・ジャパン株式会社(2005)