

感情表現に基づく小説の俯瞰分析への取り組み

吉田 知世[†] 小林 一郎^{††}

[†] お茶の水女子大学理学部情報科学科 〒112-8610 東京都文京区大塚 2-1-1

^{††} お茶の水女子大学大学院人間文化創成科学研究科理学専攻 〒112-8610 東京都文京区大塚 2-1-1

E-mail: [†]{g0720544,koba}@is.ocha.ac.jp

あらまし 近年, Web 上に存在する膨大なテキストに含まれる情報を有効活用する方法の一つとして感情抽出の研究が進められている. 本研究では対象文書として小説を用いて, 文書中に現れる感情表現に着目し, 単語に対する感情極性や感情の種類を判断する辞書を使用して文書がどのような感情の変遷に基づき構成されているか全体を俯瞰分析する. それにより, 全体の雰囲気や感情の観点から捉えた小説の選別を行う手法を提案する.

キーワード 感情分析, 小説, 感情語辞書, KeyGraph, 自己組織化マップ

An Approach to a Bird's-Eye Analysis on Novels based on Emotional Expressions

Tomoyo YOSHIDA[†] and Ichiro KOBAYASHI^{††}

[†] Department of Information Science, Faculty of Sciences, Ochanomizu University

2-1-1 Ohtsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

^{††} Advanced Sciences, Graduate School of Humanities and Sciences, Ochanomizu University

2-1-1 Ohtsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

E-mail: [†]{g0720544,koba}@is.ocha.ac.jp

Abstract In recent years, as a method to widely make use of the information contained in a huge amount of text on the Web, many studies to extract emotional information have been progressive. In our study, we focus on emotion words and phrases in novels and then do a bird's-eye analysis on them. By this, we propose a method to allow us to be able to select novels taking their atmosphere from a viewpoint of emotion.

Key words Bird's-Eye Analysis, Novel, Dictionary for Emotion Words, KeyGraph, Self-Organization Map

1. はじめに

近年, インターネットが普及し, ニュースサイトや商品レビューサイト, ブログやアンケートなど様々な電子化されたテキストが大量に存在するようになった. このようなテキスト情報を有効に利用するためのテキストマイニング技術の研究が盛んに行われている. そのひとつとして, 消費者の意見を抽出し, 分析する意見分析手法の研究が盛んになっている. その手法における抽出対象のひとつとして, 感情抽出の研究も進められている. 感情抽出ではレビュー記事やアンケートから「～が良かった」「～が悪い」といった感情表現や, どのような文構造において感情が示されているかという感情生起表現の抽出方法が研究されており, これは, ある事象や商品に対して肯定する立場か, 反対する立場か, また, 良い商品であるかどうかの判断を行う方法の一つにもなっている. 意見分析や感情抽出の対象となるテキストとして, 先に挙げたようなレビュー記事や

アンケートを対象とした研究が多く行われているが, 本研究では, 読書の対象となる機会の多い小説に対して, 感情表現の抽出を試みる. 具体的には, 多くの小説を感情の観点から分類することや, 一つの小説の中での感情の遷移を調べ, 抽出した感情表現を一望できるようにすることで, 小説全体の雰囲気や感情の流れを踏まえて読書する本を選ぶことができるようにすることを目的とする.

2. 関連研究

感情表現を収集して単語や語句ごとに感情極性値や感情表現のタグ付けを行い, それらを辞書として編纂・利用することで, 詳細な感情表現を抽出することが可能になる. そのような辞書作成に関する研究として, 高村ら [1] は, 語彙ネットワークをスピンモデルとすることで単語の感情極性値を自動計算し, 単語とその感情極性値の対応表を作成している. 伴ら [2] は, 文中の感情誘発語を用いて感情生起表現, つまり, “書き手が感

情を引き起こされた」と読み手が推測できる表現”の抽出を行っている。感情生起表現の候補表現や、感情誘発語の収集には感情表現辞典を用いている。

3. 感情表現抽出

感情抽出手法には、構文構造のパターンから感情表現を抽出するものもあるが、本研究では、語彙に対して感情のカテゴリと正負の極性値が付与された辞書を用いて感情表現を抽出する。対象文書である小説のテキストに対し、小説の進行に沿って感情表現の抽出を行う。抽出されたデータを用いた分析として、

一つの小説における感情の変遷

小説内での感情の移り変わりをグラフに表し、分析する。

一つの小説における内容と感情の関係

小説内で用いられる重要な語と感情表現の関係を分析する。

複数の小説における感情の類似判定

複数の小説について、それらが含む感情の変遷という観点から大局的に分析する。

これら3つの観点から分析を行い、感情表現に基づいた小説の俯瞰分析を試みる。

3.1 使用テキスト

本研究では前述の通り、分析対象とするテキストとして読書の対象となる機会の多い「小説」を用いる。使用するテキストは、Web上で著作権の切れた作品を無料で公開しているインターネット電子図書館「青空文庫[3]」から収集した。

表1 青空文庫の所蔵作品の属性

类目	9. 文学	8631
綱目	91. 日本文学	7969
要目	913. 小説・物語	3641
	(児童文学)	465
所蔵総数		11052

表1に青空文庫の所蔵作品の属性を示す。青空文庫に収録されている作品は11,052作品で、図書分類法の一つである日本十進分類法に基づく分類では全体のうち78%が「文学」、文学のうち92%が「日本文学」、日本文学のうち46%が「小説・物語」という分類になっている。本研究では、この日本文学の小説・物語を対象に分析を行うことにする。さらに、今回は作品中に多くの感情表現が出てくると考え、児童書として青空文庫に掲載されている作品を用いてテキストの収集を行った。所蔵されている作品の中には旧字や旧仮名で書かれた作品もあるが、それらは扱わないことにし、新字新仮名で表記されているものから88作品を用いた。

3.2 感情表現辞典

本研究では、感情語辞書に登録する感情表現の取得に“感情表現辞典[4]”を用いている。感情表現辞典は近現代作家の作品から感情表現を収録しているものであり、感情表現の用例や語句が感情の種類ごとに収録されている。感情表現辞典では感情表現を「喜、怒、哀、怖、恥、好、厭、昂、安、驚」の10種類

で分類しており、本研究では語句編^(注1)に記載されている感情表現を使用した。感情語辞書に登録した感情の種類ごとの登録数は表2のとおりである。

表2 感情表現辞典における感情表現の種類

喜	326	怒	227	哀	274	怖	195	恥	87
好	243	厭	572	昂	333	安	131	驚	159

3.3 単語感情極性対応表

感情語辞書には感情表現辞典に加えて、高村氏によって構築された“単語感情極性対応表[1]”を用いている。感情極性は、通常“positive”と“negative”の二値属性で表されるが、この対応表では語彙ネットワークを利用して-1から+1の実数値を割り当てている。具体的には、共起する語で語彙ネットワークを構築し、共起する語の属性が等しければリンクに正の重みを与え、共起する語の属性が異なればリンクに負の重みを与えている。このうち、一部のノード(語)に予め感情極性値を与えることで、残りのノードに自動的に感情極性値を付与している。この単語感情極性対応表では、感情極性がpositiveであるほど+1に近く、negativeであるほど-1に近い値が与えられており、感情極性の対象となる語には岩波国語辞書(岩波書店)の単語55,126語が使われている。

3.4 感情語辞書の構築

感情極性対応表から得られる感情極性値と感情表現辞典から得られる感情表現の種類ごとのスコアをまとめた“感情語辞書”を構築する。感情表現辞典では、同じ語句が複数の項目に現れるため、その重複をなくすようスコアをまとめている。辞書の項目は「語句、感情極性値(正)、感情極性値(負)、喜、怒、…、安、驚」の13属性によって構成されている。

3.5 処理のながれ

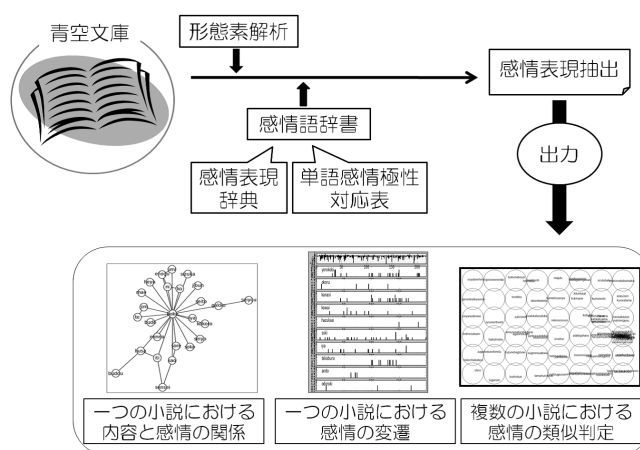


図1 処理の流れ

図1に全体の処理の流れを示す。まず、青空文庫の小説を対象テキストとして形態素解析を行う。得られた解析結果に対し

(注1): 感情表現辞典は、語句編と表現編から成っており、語句編では感情表現とされる語句そのものを、表現編では語句編に挙げられるような感情表現の用例が書かれている。

て、感情表現辞典と単語感情極性対応表を融合した感情語辞書を用いて感情表現を抽出し、感情値の算出を行う。抽出されたデータを用いて、一つの小説のデータに対しては「一つの小説における感情の変遷」および「一つの小説における内容と感情の関係」を分析し、複数の小説のデータに対しては類似度の観点から「複数の小説における感情の類似判定」を分析する。

3.5.1 テキストに対する処理

青空文庫から取得したテキストに対し、日本語形態素解析器 MeCab [5] を用いて形態素解析を行う。形態素解析から得られた語の内、品詞が動詞、名詞、形容詞、副詞であるものを抽出し、「行番号、単語の基本形」の情報を取得する。上述した4種類の品詞が、感情語辞書に登録されている語句に多くみられるものであるため、これらに限定して抽出を行った。

3.5.2 感情値の算出

形態素解析された語に対し感情語辞書を用いて、対象文書の行ごとに感情極性値と感情表現のスコアを算出する。まとめられた感情値は「行番号、感情極性値(正)、感情極性値(負)、喜、...、驚」の13属性によって構成されている(表3参照)。

表3 感情値の算出例

行番号	正	負	喜	怒	哀	怖	恥	好	厭	昂	安	驚
1	0	-0.724	0	0	0	0	0	0	0	0	0	0
2	0.24	-4.588	1	0	0	0	0	0	0	0	0	0
3	0.24	-8.06	0	0	1	2	0	1	0	0	0	0

4. 実験

感情値の算出が行われた文書に対し、感情に基づく分析を行う。一つの小説に対する分析では、「対象文書の行が進むごとに話が進んでいく」と捉え、感情の変遷をグラフに表す。さらに小説内で使われている語と感情表現との関係をグラフに表す。複数文書に対する分析では、自己組織化マップを用いて一つの小説が持つ感情の属性値に基づき、感情が類似している文書の判定を行う。

4.1 一つの小説における感情の変遷

今回、分析対象とした小説は有島武郎(1878~1923)によって書かれた「一房の葡萄」であり、この作品は「赤い鳥傑作集」に収録されている。赤い鳥傑作集は1955年に初版(新潮文庫)出版、1974年に改版(新潮社)が出版され、青空文庫の入力に使用されたものは1984年出版の44刷である。

1文ごとに語の持つ感情極性値と感情表現数を感情語辞書を用いて算出し、一房の葡萄の感情の流れを表したグラフを図2に示す。上から順に、感情極性値、「喜」、「好」、「安」、「昂」、「怒」、「哀」、「怖」、「恥」、「厭」、「驚」のグラフとなっている。感情極性値のグラフは正の極性をもつものと負の極性をもつものを分けてその極性値を表示しており、その他の感情の種類のグラフはそれらの感情表現を表す語彙の出現頻度を表している。1~5の番号が振られ、丸で囲われた部分の内容を見ると、本文中では表4の番号に対応する場面になっている。

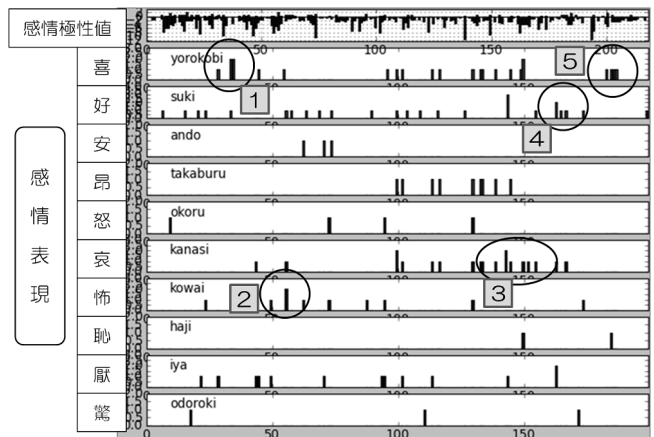


図2 「一房の葡萄」における感情の変遷に対する解析結果

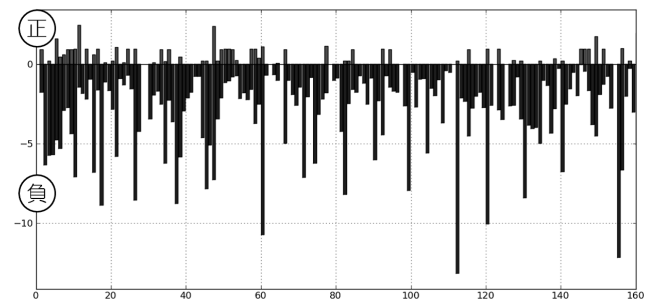


図3 感情極性値のグラフ

表4 グラフに見る感情の流れ

番号	感情の種類	本文中の内容
1	喜	お弁当の時間で楽しそうなクラスメイトの様子
2	怖	絵具を盗んだことでクラスメイトに問い詰められる
3	哀	先生に怒られそうで泣いている
4	好	クラスメイトと喧嘩をした主人公が、好きな先生に励まされる
5	喜	険悪になっていたクラスメイトと仲直りする

4.1.1 考察

グラフと本文の内容を照らし合わせると、グラフで同じ項目の感情表現が集まっている箇所では本文中でも感情を喚起する何らかの事象が多く起こっていることが分かった。グラフを通して感情表現が表れている箇所が明確になり、感情の変遷を分かりやすく提示できるようになった。

しかし、抽出された感情表現が誰の感情表現であるか判断できないといった問題や、感情表現辞典の感情の種類だけを見ても分類しにくい語句が存在するという問題がある。例えば、「驚」は「嬉しくて驚く」ことを表す表現と「悲しい知らせに驚く」ことを表す表現が存在するため、感情表現の種類とともに、感情極性値を利用して表現のされ方をより詳細に見ていく必要があると考えられる。

4.2 一つの小説における内容と感情の関係

本研究では、小説の内容の分析に大澤らによって提案されたKeyGraph [6] の一部の手法を利用する。KeyGraphは共起グラフの分割・統合操作によってキーワードを抽出する手法であ

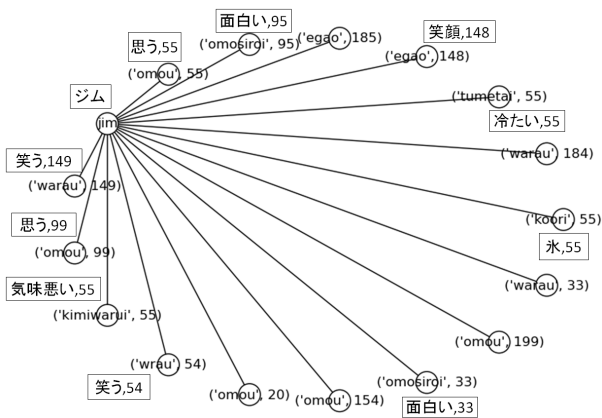


図 7 内容語(ジム)と共起する感情語と行番号のグラフ

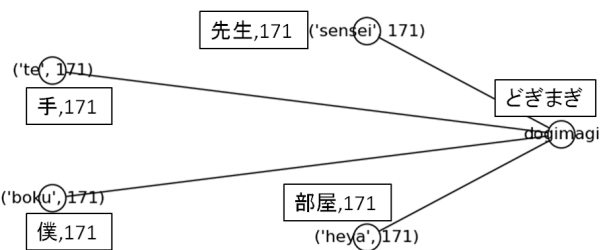


図 8 感情語(どぎまぎ)と共起する内容語と行番号のグラフ

ることができるが、図 7 では感情語に行番号が付与されることによって、文章の流れを考慮して内容語と感情語の関係を見ることができた。

図 8 は感情語“どぎまぎ”と共起している内容語とその内容語が現れる行番号がグラフに表れている。このグラフからは“どぎまぎ”という表現が出てくるのは 171 行目のみであり、その一場面で起こる出来事を内容語によって推測することができる。仮に、行番号が大きく異なった内容語が表れた場合は、感情語が表れる箇所が複数あることが分かり、行番号で複数のまとまりを作成することによって、同様に各場面で起こる出来事を推測できる。

4.3 複数の小説における感情の類似判定

自己組織化マップ(SOM)を用いて複数文書の類似判定を行う。SOMはKohonenによって提案されたニューラルネットワークの一種であり[7]、高次元データを2次元平面上へ非線形写像するデータ解析方法で、与えられた入力情報の類似度をマップ上の距離で表現するモデルである。

通常、複数の文書の類似度判定を行う場合は文書全体に対する特徴量を一括する方法を用いる。しかし、本研究では感情の変遷に関する情報も含めて類似度の判定を行うことを考え、テキストの全体を10分割し、分割されたブロックごとに12属性から成る特徴量(感情極性値(正), 感情極性値(負), “喜”, “怒”, ..., “驚”)を算出してひとつの小説の特徴量を設定している。ブロックごとにまとめられた特徴量を「タイトル, 1 ブロック目の特徴量, 2 ブロック目の特徴量, ..., 10 ブロック目の特徴量」となるようデータを作成した(図9参照)。

青空文庫の児童書のうち、88作品を対象にデータを作成し、SOMを利用して類似度の判定を行った結果を図10に示す。

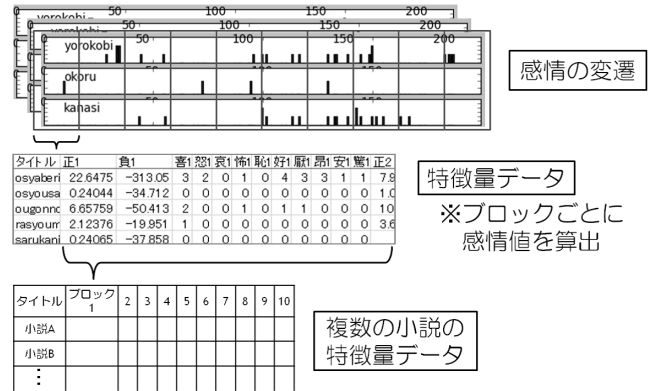


図 9 感情の変遷を考慮した特徴量の抽出

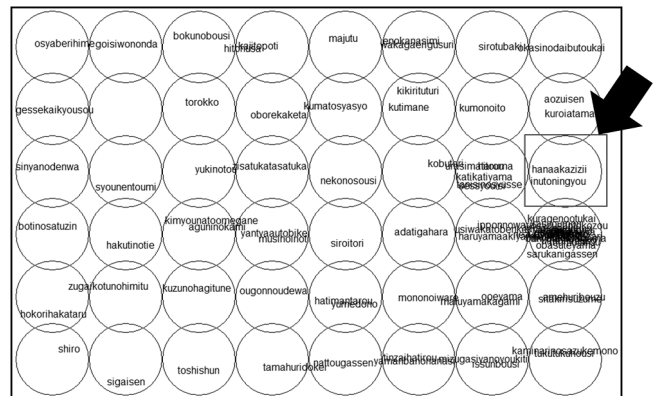


図 10 小説の類似判定結果

SOMのマッピングでは類似した小説同士が近くに配置されるため、同じ円の中に表示されたタイトルは類似していると考えられる。

4.3.1 考察

図10中の矢印で示される円の中に表示された「花咲かじじい(橋山正雄著)」と「犬と人形(夢野久作著)」という作品の感情の変遷をグラフで表した(図11参照)。グラフを見ると、“喜”, “昂”, “哀”, “厭”の4つの同じ項目で感情表現が表れていることが分かる。また、“喜”, “昂”, “哀”という3つの同じ項目においてほぼ同時に感情表現が表れている。これらの理由により、この2つの小説が類似していると判断されたと考えられる。

しかし、“喜”, “昂”, “哀”の3つの同じ項目で感情表現が表れたという点に関しては、今回はそれ以外の目立った特徴があまり表れなかったため、これらの感情表現が2つの小説の類似性が高いという判定の要因になったと考えられる。一方、同じ種類の感情表現が10分割された同じブロックにあまり現れなかったため、今回の結果からは感情の変遷を正確に追えているかは判断しにくいと考える。

5. おわりに

本研究では感情表現辞書と感情極性対応表を融合した感情語辞書を用いて、小説における感情表現を抽出した。感情に基づく小説の俯瞰分析の試みとして、一文書に対しては、感情の変遷を感情の種類によってグラフに表すことで感情が生起する部分を明確に提示することができた。さらに、内容語と感情語の

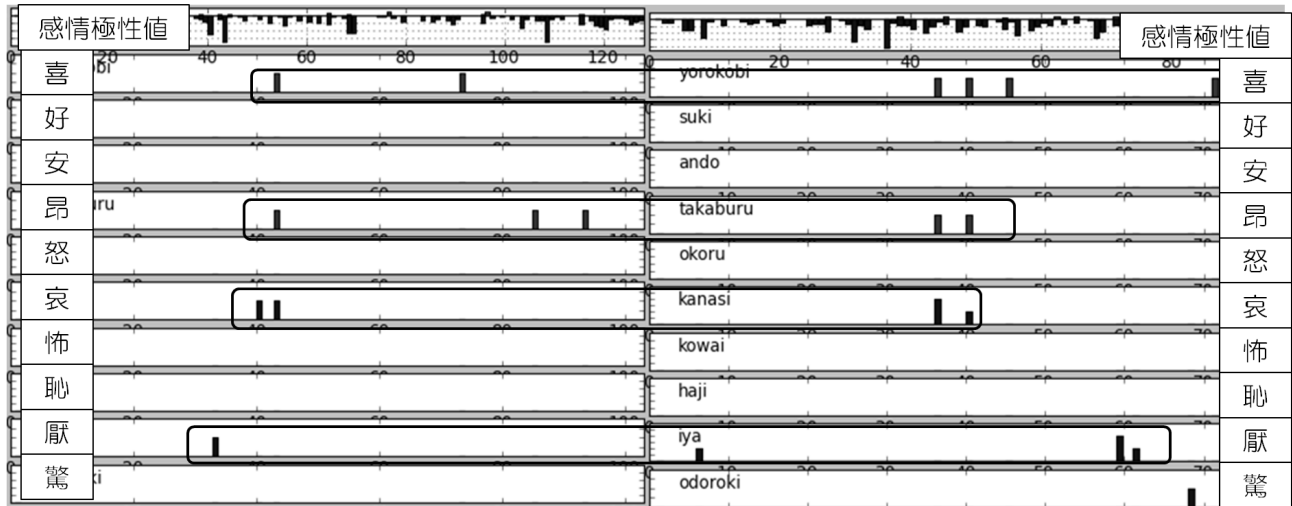


図 11 「花咲かじい」(左)と「犬と人形」(右)の比較

共起関係によってグラフを構築し、小説の主たる部分を構成する語と感情語との関係を、小説の流れを考慮しながら表すことができた。また、複数の文書に対しては SOM を利用して、類似性の観点から小説がもつ感情の分析を行った。

今後の課題として、感情極性値の使い方など感情語辞書に基づく分析手法の再検討を行う。また、内容語と感情語の関係について、さらに詳細に関係を捉えられるような場面ごとに区切る工夫や、内容語を直線上に表して話の流れに沿って感情表現を見られるようにしたいと考えている。さらに、被験者実験を行うことで提案手法の有効性を示す。

謝辞 本研究において、高村氏の構築した単語感情極性対応表を利用させて頂きました。ここに深く感謝の意を表します。

文 献

- [1] 高村大也, 乾孝司, 奥村学 “スピンモデルによる単語の感情極性抽出”, 情報処理学会論文誌ジャーナル, Vol.47 No.02 pp.627-637, 2006.
- [2] 伴昌彦, 内田理, 菊池浩明, 中西祥八郎 “係り受け関係と感情誘発語を用いた感情生起表現の抽出手法の提案”, 言語処理学会第 13 回年次大会発表論文集, pp. 404-407, 2007.
- [3] 青空文庫, <http://www.aozora.gr.jp>
- [4] 中村明 「感情表現辞典」, 東京堂出版, 1993.
- [5] MeCab, <http://mecab.sourceforge.net>
- [6] 大澤幸生, Nels E. BENSON, 谷内田正彦 “KeyGraph: 語の共起グラフの分割・統合によるキーワード抽出”, 電子情報通信学会論文誌, Vol. J82-D-I No.2 pp.391-400, 1999.
- [7] T. Kohonen, “Self-organized formation of topologically correct feature maps.”, Biological Cybernetics, 43: pp.59-69, 1982.