

形容詞との共起に基づく概念語の順序付け

仁科 俊晴^{1,a)} 吉永 直樹^{2,b)} 鍛冶 伸裕^{2,c)} 豊田 正史^{2,d)}

概要: 本稿では、概念語と形容詞の共起頻度を用いることによって、複数の概念語を、その形容詞が表す性質(「大きさ」や「高さ」など)の程度に基づいて順序付けするタスクを提起し、ウェブテキストから計算した統計量をもとにこれを解く手法を提案する。提案手法により、ある特定の状況や条件下で人々がどのような手段や物事を好むかといった行動分析や、物事に対して人々が持っている印象分析など、高度な情報収集を行うことが可能になると期待できる。頻度を求める際に利用する共起関係としては、同一文内における共起関係、係り受け関係、概念語が主格である係り受け関係の3種類を用いる。実験では、人手による順序付けとの相関を調べることにより、3種類の共起関係の順序付けにおける有用性を検証する。

1. はじめに

我々は日常生活の中で、複数の物事に対して順序関係を与えることによって情報を整理し、意思決定に利用している。例えば、外出中に食事を安くしようと思ったとき、日頃食べている料理の「食事代の安さ」を考慮して食べる料理を決めることがあるだろう。このように、関心のある性質の程度で物事を順序付けし、情報を整理することは、適切な行動を判断する上でしばしば必要であり、特に、大量の情報が存在する現代において重要であると考えられる。

本研究では複数の概念語を、ある形容詞によって表される性質(例えば「大きさ」や「高さ」など)の程度に基づいて順序付けをするタスクを提起し、大規模ウェブテキストから得た統計量によりこれを解く手法を提案する。提案手法により、ある特定の状況や条件下で人々がどのような手段や物事を好むかといった行動分析や、物事に対して人々が持っている印象分析など、高度な情報収集を行うことが可能になると期待できる。

提案手法では、与えられた概念語と、形容詞の共起頻度を概念語の持つ性質の程度の定量化に利用することで、複数の概念語を順序付けする。具体的な共起関係として、本研究では、同一文内における共起、係り受け関係、概念語が主格である係り受け関係の3種類の共起関係を検討す

る。形容詞との共起頻度を性質の程度の強さを測るのに利用する理由は、概念語が持つ性質の程度が顕著なほど、その性質に関する記述が頻繁にされるという直感に基づいており、多くの人が記述したウェブテキストから共起を数えることで、最大公約数的な順序付けが得られると考えたためである。

実験では、入力として複数の概念語とそれに共通する性質を表す形容詞を与え、テキストから得た共起頻度をもとに概念語の順序付けを行った。提案手法の有用性については、4人の被験者による概念語の順序付けとの順位相関係数を用いることで確認をした。

本稿の流れは以下の通りである。まず2章で本研究で取り組むタスクについて説明する。次に、3章で提案手法について述べ、4章ではその評価を行う。そして、5章で関連研究について述べる。最後に、6章でまとめと今後の課題について述べる。

2. 概念語の順序付けタスク

本章では、本研究で取り組む概念語の順序付けというタスクの定義を具体例を用いて説明する。

本研究では、例えば、「カレー」、「寿司」、「ステーキ」のような複数の概念語と、それらに共通する性質を表す形容詞(例えば「安い」)を入力として与え、入力された概念語を指定された性質の程度(ここでは「安さ」)によって順序付けするタスクを考える。概念語が持つ性質の種類はある程度限られているため、複数の概念語と形容詞のペアは入力として与えられることを仮定した上で、順序付けを行う。なお、概念語には、与えられた形容詞(性質)において相互に順序付け可能な名詞(句)を考える。

¹ 東京大学大学院 情報理工学系研究科
Graduate School of Information Science and Technology,
The University of Tokyo

² 東京大学 生産技術研究所
Institute of Industrial Science, The University of Tokyo

a) nishina@tkl.iis.u-tokyo.ac.jp

b) ynaga@tkl.iis.u-tokyo.ac.jp

c) kaji@tkl.iis.u-tokyo.ac.jp

d) toyoda@tkl.iis.u-tokyo.ac.jp

文	共起関係		
	同一文内共起	係り受け	主格係り受け
安く昼食を済ませるために、カレーを食べた	✓		
いつもより安いカレーを買った	✓		
カレーって安いよね	✓	✓	
カレーよりも安いそばを食べた	✓	✓	
カレーは安いから最高だ	✓	✓	✓
カレー安くて助かる	✓	✓	✓
カレー、本当に安い	✓	✓	✓

表 1 共起頻度の集計対象となる文

概念語を順序付けする際には、(1) 概念語の指示対象が曖昧であること、(2) 程度の判断に主観が含まれることから、絶対的に正しい順序付けを定義することは難しい。(1) の場合では、例えば、食べ物を安さでランキングしようとするときに高いカレーもあれば安いカレーもあるように、どの「カレー」を想定するかによって順位が逆転する。しかしながら、我々は通常は物事を典型的事例によって認識しており [2]、そのような認識に従えば、例えば寿司がカレーより高いというような順序付けは、異なる評価者の間でもある程度の一致が得られると期待できる。一方で、(2) の場合、例えば食べ物を美味しさで順序付けするような場合については、個人による好みの違いがあるため、順序付け結果が一致しないことが予想される。

本研究の目的は、(1) や (2) の問題が存在する概念の順序関係について人々の共通認識を導けるかという問いに答えることにある。

順序付けする対象によってどれだけ個人の順序付けに揺れが生じるかについては、4 人の被験者の順序付けの相関係数を求めることによって確認する。

3. 提案手法

我々は、各概念語について、その程度が顕著な性質は頻繁に言及される (例えば、「クジラは大きい」という言明は「ネズミは大きい」という言明より、相対的に高頻度で書かれる) という直感に基づき、概念語と形容詞の共起頻度によって概念語が持つ性質の程度の強さをスコア化することを提案する。

以下、本章では 2 章で定義した概念語の順序付けタスクをウェブテキストから得られる形容詞との共起頻度により概念語をスコア付けすることで解く手法を説明する。また、共起頻度の計算方法として幾つかの方法を提案し、それらの比較を行う。

3.1 順序付けの手がかりとする概念語と形容詞の共起関係

本研究では、手がかりの頻度を求める際に利用する概念語と形容詞の共起関係として、同一文内における共起関係、係り受け関係、概念語が主格である係り受け関係の 3 種類を考える。同一文内における共起と係り受け関係を比較す

ると、係り受け関係では形容詞が概念語の性質を記述することが保証されているため、性質の程度の強さのスコア付けに用いる共起としてはより適切であると考えられる。しかしながら一方で、係り受け関係は同一文内における共起関係に比べて頻度が少なくなるため、クエリによってはスコア付けに十分な統計量が得られない可能性がある。このように、これらの共起は質・量の観点でトレードオフの関係にあるため、本研究では、それら 3 種類の共起頻度計算方法を比較することにより、その優劣を検証する。

以下では、概念語「カレー」と形容詞「安い」を例として、共起関係を説明をする。

同一文内共起

同一文内で概念語と形容詞が同時に出現した頻度。「スーパーでカレーの特売がやっていたので、非常に安い値段でカレーが作れました。」という文の場合は (組み合わせにより) 共起が 2 回、「カレーって安いけど、カレーより安い食べ物もたくさんあるよね。」という文の場合は (組み合わせにより) 共起が 4 回観測されたとする。

係り受け

概念語から形容詞への係り受け関係の頻度。例えば、「カレーって安いよね。」などの文から共起頻度を求める。前述の同一文内共起では、「カレーを食べた後行った喫茶店のコーヒーは安かった」といった文も共起頻度の集計対象になるが、この文は、カレーの安さについて言及しているわけではなく、本論文でスコア付け算出に用いるには適切でない場合がある。係り受け関係を手がかりとすることにより、このような不適切な共起を除外することができる。

主格係り受け

概念語から形容詞への係り受け関係で概念語が主格であるものの頻度。主格の判定には、概念語に付属する助詞を用いる。具体的には、概念語と同一文節内に格助詞「は」「が」「も」のいずれかが出現している場合、もしくは格助詞が全く出現していない場合、その概念語が主格であると判断する。係り受けではあるが、主格係り受けではなく、本論文での共起頻度の集計対象として扱うには不適切な文の例として「カレーより安

形容詞	概念語集合
大きい	ネコ, ウマ, クマ, ウシ, ネズミ, クジラ, キリン, イヌ, ゾウ, サル
速い	自動車, 自転車, 飛行機, 電車, ヘリコプター, スクーター, 新幹線, 船
重い	タンス, じゅうたん, ソファ, 椅子, ベッド, 電気スタンド, 机, カーテン
重い	ストーブ, ホットプレート, コーヒーメーカー, 冷蔵庫, アイロン, 掃除機, 電子レンジ, トースター, 洗濯機
安い	焼きそば, ハンバーガー, チャーハン, パン, 寿司, カレー, ピザ, ステーキ, パスタ
甘い	ナシ, マンゴー, リンゴ, ミカン, モモ, レモン, カキ, イチゴ, グレープフルーツ, パイナップル
可愛い	ウサギ, リス, カメ, ウマ, ネコ, イヌ, ヒツジ, トカゲ, ハムスター, サル
美味しい	焼きそば, ハンバーガー, チャーハン, パン, 寿司, カレー, ピザ, ステーキ, パスタ

表 2 形容詞と概念語集合の組み合わせ

いそばを食べた。」のような文がある。

なお、係り受け関係のうち、概念語から形容詞への場合のみを利用する理由は、「いつもより**安いカレー**を買った。」といった限定修飾を共起頻度の集計対象から除外するためである。上記のような文は、ある特定の**カレー**が**安い**ことに関する言及であり、**カレー**が**安い**食べ物であることを述べているわけではない。このような文から共起頻度を計算してしまうと、不自然な順序付けを導く可能性がある。

表 1 に概念語が「**カレー**」、性質を表す形容詞が「**安い**」の場合に、3 種類の各共起関係が現れる文の例を示す。なお、本研究では、概念語と形容詞の共起頻度によって性質の程度を定量化するため、形容詞の直後(同一文節中)に否定形が続く文は共起頻度を計算する対象から除外した。

3.2 共起に基づく概念語のスコア付け

3.1 節で説明した共起頻度をもとに、各概念語自体の出現頻度を考慮してスコア付けする。

$$\text{順序付けのスコア} = \frac{\text{概念語と形容詞の共起頻度}}{\text{概念語の出現頻度}}$$

概念語の出現頻度をスコア付けで考慮する理由は、概念語と形容詞との共起頻度を正規化し、同じ尺度で順序付けを行うためである。このようにして得られたスコアをもとに降順で概念語を並べることによって順序付けを行う。

4. 評価実験

本章では、前章で述べた提案手法を用いて複数の概念語の順序付けを行う。さらに、人手で順序付けされた複数の概念語との順位相関係数を求めることにより、順序付けに用いる 3 種類の共起関係の優劣を検証する。

共起頻度の計算に用いるウェブテキストは、2006 年 2 月から 2012 年 7 月までに書かれたブログ記事(約 2 億記事, 19 億文)を利用した。また、形態素解析には Kaji らの手法 [3], 係り受け解析には Yoshinaga らが開発した J.DepP[4][5] を利用した。

4.1 評価用データの作成

実験に先立ち、評価用データを作成した。入力として与

えられる概念語集合と形容詞のペアは、順序付けをするのに適した組み合わせでなくてはならない。そこで、実験の評価に用いる概念語集合と形容詞の組み合わせを以下のよう手順で生成した。

まず、入力に用いる形容詞をテキスト中での頻度を基準に選択した。具体的には、ウェブテキスト中で頻度上位の形容詞から「ない」「普通だ」「すごい」といった、性質の程度を表現しない形容詞を除いて選択した。

次に、このようにして得た各形容詞に対して、その形容詞に高頻度で係っていた名詞(句)の中から概念語を 1 つ選んだ。なお、概念語を選ぶ際に、その語と比較可能な概念語を想起しにくい場合は、その概念語は候補から除外した。例えば、「大きい」に対する概念語を選択する際には、候補としては「声」「胸」「ネコ」などがあったが、「声」や「胸」は順序付けが可能な概念語集合を想起しづらかったため、候補から除外した。結果、頻度順位が下であっても、「イヌ」や「ウサギ」などの「大きさ」を比較可能な類似概念語が存在する「ネコ」を選ぶこととなった。

このようにして得られた形容詞と概念語のペアに対し、言語資源を利用して類似する概念語を収集することで入力の形容詞と概念語集合を得た。本研究では、日本語 WordNet^{*1}と NLTK (Natural Language Toolkit)^{*2}の WordNet モジュール [6] を用いて類似する概念語を最大 10 語列挙した。

以上の手続きにより、最終的な評価用データとなる概念語集合と形容詞のペアを 8 セットを用意した。具体的な概念語集合と形容詞の組み合わせは表 2 に示す。

人手による概念語集合の順序付けは 4 人の被験者の回答をもとに構築した。

4.2 評価尺度

提案手法によって得られた順位と人手によって作成された順序付けの相関を、スピアマンの順位相関係数を用いて評価した。本研究におけるスピアマンの順位相関係数 ρ は、順序付けをする概念語の総数を N 、順位相関を求める 2 つの順位を X と Y 、 X と Y における概念語 i の順位差

*1 <http://nlpwww.nict.go.jp/wn-ja/>

*2 <http://nltk.org/>

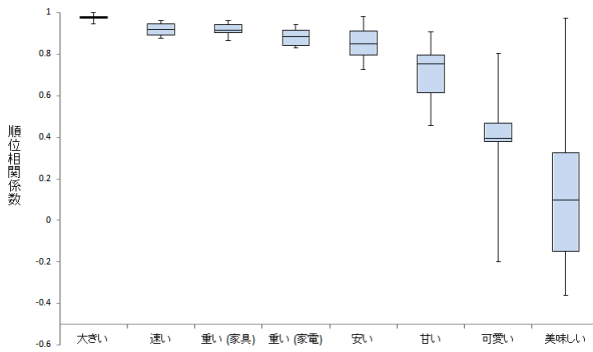


図 1 被験者間の順序付けの一致度

を D_i , X と Y において同順位の概念語の個数をそれぞれ n_x と n_y としたときのそれらの順位を t_i と t_j ($i=1,2,\dots,n_x$: $j=1,2,\dots,n_y$) としたとき、以下のように計算する。

$$\rho = \frac{T_x + T_y - \sum_{i=1}^N D_i^2}{2\sqrt{T_x T_y}}$$

$$T_x = \frac{N^3 - N - \sum_{i=1}^{n_x} (t_i^3 - t_i)}{12}$$

$$T_y = \frac{N^3 - N - \sum_{i=1}^{n_y} (t_j^3 - t_j)}{12}$$

順位相関係数は $-1 \leq \rho < 1$ までの値をとる。 $0 \leq |\rho| < 0.2$ はほぼ相関なし、 $0.2 \leq |\rho| < 0.4$ は低い相関あり、 $0.4 \leq |\rho| < 0.7$ は相関あり、 $0.7 \leq |\rho| < 1$ は高い相関あり、と評価される。 ρ が負の場合は負の相関について同様のことがいえる。

なお、スピアマンの順位相関係数は 2 つの順位の相関を求めるものであるため、被験者が与えた順序付けとの相関は、各被験者による順序付けとの順位相関係数の平均を算出することで評価した。

4.3 各順序付け対象における被験者間の一致度

与えられた各概念語集合の順序付けにおいて、人々の間でどれだけ一致度が異なるかを計るために、被験者同士の順序付けの間の相関を評価した。図 1 に、各入力に対する 4 人の被験者の順序付け結果の順位相関係数を示す。また、表 3 に、被験者の順位相関係数の平均を示す。「可愛い」、「美味しい」に基づく概念語の順序付けに関しては、主観による揺れが大きかったが、それ以外の性質に関する順序付けに関しては、順序付けが高い相関を持っていることが確認された。図 1 から分かるように、多くの入力について、被験者間で概念の順序関係に関する共通認識が確認されたため、平均的な概念語集合の順序付けが有用であることが分かった。

4.4 実験結果

表 4 に被験者が付与した順序付けと提案手法が付与した順序付けとの間の順位相関係数を示す。また、全てのありうる順序付けを考慮することで計算した順位相関係数の平

形容詞	順位相関係数
大きい	0.977
速い	0.921
重い(家具)	0.918
重い(家電)	0.884
安い	0.854
甘い	0.709
可愛い	0.377
美味しい	0.163

表 3 被験者の付けた順位の平均順位相関係数

形容詞	上限値	共起関係		
		同一文内共起	係り受け	主格係り受け
大きい	0.983	0.438	0.462	0.558
速い	0.949	-0.435	-0.164	-0.152
重い(家具)	0.958	0.662	0.217	0.259
重い(家電)	0.939	0.216	0.209	0.209
安い	0.924	-0.468	-0.535	-0.535
甘い	0.866	0.121	0.182	0.103
可愛い	0.722	0.439	0.428	0.418
美味しい	0.598	-0.110	-0.282	-0.282
平均	0.867	0.108	0.0632	0.0753

表 4 各形容詞に基づいて順序付けされた相関係数とその平均

均値の上限値も併せて示す。表から確認できるように、「速い」、「安い」、「美味しい」以外の入力については正の相関が見られた。

表 5 から表 12 に、各共起関係によって実際に順序付けされた概念語とその順位を示す。表 5 の「大きい」に基づく順序付けについて、高い順位相関係数を得られた共起関係は順に、主格係り受け、係り受け、同一文内共起であった。「大きい」に関しては、主格係り受けで共起を考慮した場合でも十分な頻度で各概念語と共起しており、共起関係をより厳密に考えることで良い順序付けが得られるという、直感に合う結果が得られた。一方、他の形容詞に基づいた順序付けの結果に関しては、予想とは異なる結果となった。これについて、次節で概念語の順位が共起関係によってどのように変わるかを分析し、その原因を考察する。

4.5 順序付け誤りの考察

前節の実験において、順序付けが不自然と思われる概念語についてその要因を調べた。順序付けの誤りとして、大きく分けて (a) 想定される順位よりも高い順位となる (過大評価) (b) 想定される順位よりも低い順位となる (過小評価) の 2 つの場合に分けられる。本章では、それぞれの場合について、実際に見られた具体的な例を示すとともに順序付け誤りの原因を分析する。

a) 想定される順位よりも高い順位の場合 (過大評価)

この誤りに当てはまる概念語とその性質は、「ネズミ」の「大きさ」、「スクーター」の「速さ」、「パイナップル」「イチゴ」「ミカン」の「甘さ」、「アイロン」「掃除

機」「椅子」の「重さ」,「ステーキ」「寿司」の「安さ」などが挙げられる。これらは3種類の共起関係全てにおいて過大評価となっていた。これらについては、最も厳密な主格係り受けを共起に用いても過大評価となっていたため、主格係り受けの共起が、どのような文に存在していたのかを調べた。以下に、顕著に見られた文を類別して報告する。

意外性のある事象を取り上げて報告する文

ブログ記事では、ユーザは自身が意外な発見をした時にその発見を記事にすることがある。具体的な例を以下に示す。

- 先日購入したスクーターは想像以上に速かった。
- ネズミが思いのほか大きくて驚いた。

このような文が共起頻度の計算対象に含まれた結果、スクーターは速い乗り物、ネズミは大きい動物という不自然な結果が得られたと考えられる。

限定的な文脈で成立する事象を記述した文

また、特定の条件の下でのみ成立する事象を記述した文も多く見られた。具体的な例を以下に示す。

- 都会のネズミは大きいんだって。
- ウチのアイロンは重いので、すげえ掛けづらかったです。
- 今のスクーターは速い。
- ここは、ステーキが安くて美味しい。

例えば最初の文は、「本来のネズミは大きくないが、都会のネズミに限り大きい」という意味であり、ネズミの(普遍的な)大きさを記述した文ではない。これが今回の提案手法で負の順位相関係数になった順序付け結果が得られた理由のひとつであると考えられる。

形容詞の多義性

同じ形容詞に複数の意味があることが原因で概念語と形容詞の共起が多く観測された文も多く見られた。以下に具体的な例を示す。

- 店で見たときにはいい感じだったけど、いざ家に置いてみると椅子は少し重い感じがしました。
- この椅子は重い感じの色合いだなー。

これらの「重い」は重量ではなく雰囲気のことを言っている。このような共起が多く観測された結果、椅子は重い家具、という不自然な結果が得られたと考えられる。

b) 想定される順位よりも低い順位となる(過小評価)

この誤りに当てはまる概念語とその性質は、「ベッド」の「重さ」、「ヘリコプター」の「速さ」、「モモ」の「甘さ」、「カレー」「パン」の「安さ」などが挙げられる。これらも(a)と同様、3種類の共起関係全てにおいて過小評価となっていた。過小評価の理由は概念語と形容詞との間に共起関係があまり観測されなかったため

である。考えられる原因を以下に列挙する。

概念語が持つ性質の多様性

概念語と形容詞の共起頻度を概念語自体の出現頻度で正規化することでスコア付けを行っているため、概念語が様々な性質と共起する場合、個々の形容詞に対して共起頻度が分散し、(それぞれの性質の程度が顕著であったとしても)スコアが低くなることがある。例えば、「カレー」は「安さ」「美味しさ」の他に「辛さ」などの性質も話題に上がりやすく、正規化によりスコアが過小評価されたと考えられる。

概念語の多義性

概念語が多義である場合、概念語の頻度による正規化によってスコアが過小評価されることがある。例えば、「モモ」の「甘さ」がこの場合に当てはまる。「モモ」に関する文を見てみると、果物の「モモ」だけでなく、肉の「モモ」や人名の「モモ」などに対する文も含まれていた。これによって、概念語の出現頻度が多くなり、順序付けのスコアが低くなる。その結果、過小評価につながったと考えられる。

当然の事象に関する記述不足

例えば、ベッドが重いものであったり、パンが安いものであったりすることは、一般的には当然のことである。当然の事象は実際に文として記述される機会が少ないのでその結果、順序付けのスコアが低くなり、過小評価につながったと考えられる。これは、(b)の(1)と逆の場合である。

5. 関連研究

著者の知る限り、我々と同種の順序付けタスクに取り組んだ研究はこれまで存在しないが、物事や概念の間に順序付けを行おうという試みはいくつか存在している。以下でそれらの研究について簡単に紹介し、本研究との違いを述べる。

倉島らは具体物を「良さ」によって順序付けするタスクに取り組んでいる[1]。この研究では、比較情報をもとに有向グラフを作成し、各ノードの評価値を求めることにより順序をつけている。彼らが想定する順序付けの尺度は「良さ」だけであり、我々のタスクは彼らのタスクを一般化したものになっている。

形容詞の極性(肯定・否定)や、記述する程度の強さを獲得する研究も行われている。Turneyは形容詞が肯定的、あるいは否定的な形容詞かどうかを、その程度も含めてウェブテキストから得られた統計量をもとに計算する手法を提案している[7]。また、de MeloとBansalは同種の性質(例:暑さ)を記述する形容詞(例:warm, hot, scorching)を、表現する程度の強さで順序付けする手法を提案している[8]。これらの研究は形容表現自体の程度の強さを順序付けしているものであり、概念語をその性質の程度に基づ

いて順序付けする本研究とは異なる。

6. おわりに

本研究では、類似する物事の順序付けが情報分析や高度な情報収集を実現をする上で重要であるということに着目し、複数の概念語を共通する性質の程度に基づいて順序付けするタスクを提起した。また、ウェブテキストから得られる統計量を用いて入力概念語と形容詞の共起を算出し、順序付けの指針となるスコアを得る手法を提案した。

今後、より精度の高い順序付けをするためのアプローチとして以下の2つを考えている。

1つ目は、形容詞以外の、概念語の性質の程度を表す記述との共起関係を利用することである。具体的には、比較表現、強調表現、反義形容詞との間の共起関係を利用することを検討している。

比較表現とは、「～よりも」や「～に比べて」などの表現である。比較表現を考慮することにより、不自然な順序付けの原因となった、一般的には成立しない事象に関する問題を緩和できるのではないかと考えている。

強調表現とは、「とても」や「かなり」といった修飾語や、「案外」や「意外」といった単語である。このような語を伴う記述は、そうでない記述よりもスコア付けを行う際に重視すべきである。

反義形容詞とは、形容詞「安い」に対する形容詞「高い」などである。例えば、今回の実験で「寿司」は入力として与えた概念語の中でも「安い」部類であるという結果になった。反義形容詞の共起度も考えることにより、「寿司」の「高さ」に言及した文もスコア付けする際の集計対象として扱えるので、より頑健な順序付けができるようになると思う。

2つ目は、順序付けのスコア化について再考することである。本研究では、3種類の共起関係全てにおいて、概念語の出現頻度によって正規化をした。これについて、例えば、概念語と形容詞の係り受け関係を手がかりにする場合は、概念語が係り受け関係を持つ文の頻度で正規化するなど、共起関係に応じて正規化を変えた実験も行うべきである。

また、テキストデータを今回使用したブログ記事からマイクロブログに変えて同実験を行い、結果を比較することで情報源の違いや量による精度の差異についての議論も検討する。

謝辞

本研究は JSPS 科研費 25280111 の助成を受けたものである。

参考文献

[1] 倉島健, 別所克人, 戸田浩之, 内山俊郎, 片岡良治, 奥雅博. 比較情報に基づくランキング手法. *DBSJ*, Vol. 6, No. 1,

2007.

- [2] John R. Taylor. *Linguistic Categorization*. Oxford: Clarendon Press, 1989.
- [3] Nobuhiro Kaji and Masaru Kitsuregawa. Efficient Word Lattice Generation for Joint Word Segmentation and POS tagging in Japanese. *In Proceedings of IJCNLP*, October 2013.
- [4] Naoki Yoshinaga and Masaru Kitsuregawa. Kernel Slicing: Scalable Online Training with Conjunctive Features. *In Proceedings of COLING*, pp. 1245–1253, 2010.
- [5] Naoki Yoshinaga and Masaru Kitsuregawa. Polynomial to Linear: Efficient Classification with Conjunctive Features. *In Proceedings of EMNLP*; pp. 1542–1551, 2009.
- [6] Dekang Lin. An Information-Theoretic Definition of Similarity. *ICML*, pp. 296–304, 1998.
- [7] Peter D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *In Proceedings of ACL*, pp. 417–424, July 2002.
- [8] Gerard de Melo and Mohit Bansal. Good, Great, Excellent: Global Inference of Semantic Intensities. *In Proceedings of ACL*, pp. 279–290, 2013.

順位	同一文内共起			係り受け			主格係り受け		
	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度
1	クジラ	942	49592	ゾウ	149	55981	ゾウ	38	55981
2	ゾウ	993	55981	クジラ	111	49592	クジラ	26	49592
3	ネズミ	1360	99178	ネズミ	159	99178	キリン	35	87750
4	ウシ	144	14069	キリン	121	87750	ネズミ	32	99178
5	キリン	884	87750	クマ	177	177801	クマ	46	177801
6	クマ	1763	177801	ネコ	260	295009	ネコ	65	295009
7	ネコ	2425	295009	ウン	12	14069	ウシ	3	14069
8	サル	628	89388	イヌ	19	42775	ウマ	6	28762
9	イヌ	287	42775	ウマ	12	28762	サル	17	89388
10	ウマ	150	28762	サル	35	89388	イヌ	6	42775
ρ		0.438			0.462			0.578	

表 5 形容詞「大きい」に基づく順序付け

順位	同一文内共起			係り受け			主格係り受け		
	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度
1	新幹線	2815	548589	新幹線	1052	548589	新幹線	536	548589
2	スクーター	175	36913	スクーター	30	36913	スクーター	10	36913
3	船	1752	643238	飛行機	334	726960	飛行機	159	726960
4	自転車	3642	1360401	船	254	643238	船	134	643238
5	自動車	449	198698	自転車	519	1360401	自転車	254	1360401
6	飛行機	1366	726960	自動車	66	198698	電車	210	1850311
7	電車	2533	1850311	電車	443	1850311	自動車	12	198698
8	ヘリコプター	68	49804	ヘリコプター	10	49804	ヘリコプター	1	49804
ρ		-0.435			-0.164			-0.152	

表 6 形容詞「速い」に基づく順序付け

順位	同一文内共起			係り受け			主格係り受け		
	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度
1	タンス	315	53907	タンス	29	53907	タンス	16	53907
2	机	1120	399452	椅子	86	449851	椅子	54	449851
3	椅子	1171	449851	じゅうたん	4	21598	机	39	399452
4	ベッド	1425	561214	机	70	399452	じゅうたん	2	21598
5	じゅうたん	53	21598	カーテン	30	212034	カーテン	13	212034
6	ソファ	363	152963	ソファ	17	152963	ソファ	9	152963
7	カーテン	471	212034	ベッド	54	561214	ベッド	30	561214
8	電気スタンド	6	3645	電気スタンド	0	3645	電気スタンド	0	3645
ρ		0.662			0.217			0.259	

表 7 形容詞「重い」に基づく順序付け (家具)

順位	同一文内共起			係り受け			主格係り受け		
	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度
1	掃除機	853	147514	掃除機	129	147514	掃除機	96	147514
2	アイロン	211	65662	アイロン	35	65662	アイロン	24	65662
3	洗濯機	412	168768	洗濯機	46	168768	洗濯機	33	168768
4	ホットプレート	35	20113	ホットプレート	4	20113	ホットプレート	3	20113
5	冷蔵庫	776	569723	電子レンジ	12	94247	電子レンジ	5	94247
6	ストーブ	168	130863	冷蔵庫	55	569723	冷蔵庫	28	569723
7	電子レンジ	80	94247	ストーブ	7	130863	ストーブ	6	130863
8	コーヒーメーカー	9	15952	コーヒーメーカー	0	15952	コーヒーメーカー	0	15952
9	トースター	14	27630	トースター	0	27630	トースター	0	27630
ρ		0.216			0.209			0.209	

表 8 形容詞「重い」に基づく順序付け (家電)

順位	同一文内共起			係り受け			主格係り受け		
	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度
1	ステーキ	1214	120860	ステーキ	114	120860	ステーキ	79	120860
2	寿司	2888	334108	寿司	259	334108	寿司	171	334108
3	ハンバーガー	877	117581	ハンバーガー	82	117581	ハンバーガー	48	117581
4	焼きそば	714	131127	ピザ	149	259153	ピザ	97	259153
5	ピザ	1384	259153	焼きそば	71	131127	焼きそば	49	131127
6	チャーハン	472	99093	チャーハン	41	99093	チャーハン	25	99093
7	パスタ	2264	512502	パスタ	194	512502	パスタ	119	512502
8	パン	5265	1575387	パン	478	1575387	パン	318	1575387
9	カレー	2820	856243	カレー	229	856243	カレー	133	856243
ρ		-0.468			-0.535			-0.535	

表 9 形容詞「安い」に基づく順序付け

順位	同一文内共起			係り受け			主格係り受け		
	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度
1	マンゴー	5442	116503	マンゴー	885	116503	イチゴ	563	256116
2	イチゴ	11598	256116	パイナップル	389	61490	ミカン	108	55928
3	パイナップル	2460	61490	イチゴ	1381	256116	パイナップル	113	61490
4	グレープフルーツ	1588	54996	リンゴ	699	194695	マンゴー	189	116503
5	リンゴ	5417	194695	ミカン	189	55928	リンゴ	161	194695
6	ミカン	1541	55928	グレープフルーツ	124	54996	グレープフルーツ	36	54996
7	レモン	4404	197623	レモン	186	197623	レモン	92	197623
8	ナシ	447	106836	カキ	25	42554	カキ	7	42554
9	カキ	172	42554	ナシ	40	106836	ナシ	14	106836
10	モモ	483	175111	モモ	47	175111	モモ	10	175111
ρ		0.121			0.182			0.103	

表 10 形容詞「甘い」に基づく順序付け

順位	同一文内共起			係り受け			主格係り受け		
	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度
1	ウサギ	5405	165691	リス	365	63668	リス	191	63668
2	リス	2067	63668	ハムスター	203	39451	ハムスター	104	39451
3	ハムスター	1060	39451	ウサギ	756	165691	ウサギ	354	165691
4	ネコ	5894	295009	ネコ	966	295009	ネコ	462	295009
5	ヒツジ	257	14566	ヒツジ	39	14566	ヒツジ	18	14566
6	カメ	1407	106138	トカゲ	83	34925	トカゲ	39	34925
7	トカゲ	462	34925	カメ	247	106138	カメ	93	106138
8	サル	921	89388	サル	145	89388	サル	68	89388
9	イヌ	385	42775	イヌ	50	42775	イヌ	30	42775
10	ウマ	79	28762	ウマ	5	28762	ウマ	3	28762
ρ		0.439			0.418			0.418	

表 11 形容詞「可愛い」に基づく順序付け

順位	同一文内共起			係り受け			主格係り受け		
	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度	概念語	共起頻度	概念語頻度
1	パスタ	38668	512502	ピザ	4527	259153	ピザ	3620	259153
2	ピザ	19144	259153	チャーハン	1616	99093	チャーハン	1272	99093
3	パン	112609	1575387	パスタ	7559	512502	パスタ	5986	512502
4	ステーキ	8065	120860	パン	19661	1575387	パン	15439	1575387
5	チャーハン	6510	99093	ステーキ	1492	120860	ステーキ	1127	120860
6	寿司	20241	334108	カレー	10241	856243	カレー	7673	856243
7	カレー	47838	856243	ハンバーガー	1359	117581	ハンバーガー	998	117581
8	ハンバーガー	6102	117581	寿司	3421	334108	寿司	2617	334108
9	焼きそば	6080	131127	焼きそば	1246	131127	焼きそば	900	131127
ρ		-0.110			-0.282			-0.282	

表 12 形容詞「美味しい」に基づく順序付け