

修 士 論 文

大規模ウェブテキストを用いた
多様な観点に基づく概念語の順序付け

Ordering Concepts from
Various Viewpoints
by using Large-scale Web Text

指導教員 豊田 正史 准教授



東京大学大学院情報理工学系研究科
電子情報学専攻

氏 名 48-126433 仁科 俊晴

提 出 日 平成26年2月6日

概 要

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。この時、多くの料理の候補を列挙したり「安さ」以外にも「美味しさ」や「近さ」など多くの観点を考慮したりすることでより良い選択が可能になるが、全てを予め列挙することは難しい。

以上のことから、計算機により自動的に物事の順序付けを行うことは有用であると考えられるが、その実現のためには2つの課題が存在する。1つ目は順序付けの対象とする概念語や順序付けする際の観点をどのように列挙するかということであり、2つ目は実際にどのように順序付を行うかということである。本論文では、これら2つの課題を概念語集合からそれらを順序付けの観点となる形容詞を獲得するタスクと複数の概念語を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に分け、それを実現するための方法を提案する。本研究では、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）を与えるものとする。

評価実験において、順序付けの観点となる形容詞の獲得タスクにおいては入力として与えた概念語集合からスコア化した上位10位の形容詞の適合率を4人の被験者によって確かめることで提案手法の有用性を示した。概念語の順序付けタスクにおいては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提案手法の有用性を示した。

謝辞

はじめに，指導教官である豊田准教授に深く感謝いたします．豊田准教授には本研究を進めるにあたって貴重な意見を何度もしていただきました．また，計算機のトラブルがあった時やなにか分からないことがあった時に質問をすると常にすぐ回答をしてくださり大変助かりました．

次に，鍛冶特任准教授と吉永准教授に深く感謝を致します．お忙しいにも関わらず私が質問や進捗報告をしに伺えば常に親身になって対応をして下さり，いつもの確なアドバイスをしてくださいました．自然言語処理の分野に関して十分な知識を持ち合わせないため，稚拙な質問をすることも多々ありましたが，それでも毎回親切に答えてくださったことを感謝致します．鍛冶特任准教授と吉永准教授の助けがなければ満足いく研究生活が送れてはいなかったと言っても過言ではありません．また，論文の添削や発表練習に関しては何度も付き合ってください感謝しております．

次に，修士ミーティングや自然言語処理ミーティングに参加してくださった横山大作助教と伊藤正彦助教に深く感謝を致します．私がミーティングで発表をすると横山助教からは毎回するどい質問が来るので，研究の本質を見直し，研究のゴールを再確認する良い機会となっていました．また，伊藤助教には研究でのご指導だけでなく，研究生活でも大変お世話になりました．研究室での行事の際には取りまとめを行って下さりありがとうございました．より良い研究生活が送れたのも伊藤助教のおかげです．

併せて，理想的な研究環境を提供してくださった喜連川優教授，中野美由紀特任准教授をはじめとするその他の研究室スタッフの方々，研究生活のサポートをしてくださった秘書の方々，忙しいにも関わらず早急に大量実験データを評価してくれ

た後輩にも感謝します。

最後に、研究室での宿泊が長くなると私のことを気にかけて連絡をしてきてくれた家族へ感謝します。

2014年2月6日

目次

謝辞	i
第1章 はじめに	1
1.1 ウェブテキストを用いた順序付け知識獲得の重要性	1
1.2 本研究の目的と貢献	2
1.3 本論文の構成	3
第2章 アプローチ	4
2.1 共起頻度獲得における表記揺れの統合	4
2.2 概念語と形容詞の共起関係の獲得	5
2.3 順序付け対象となる概念語集合の獲得	6
第3章 順序付け観点の形容詞獲得	9
3.1 提案手法	9
3.1.1 比較文を用いたフィルタリング	10
3.1.2 共起頻度に基づくスコアリング	10
3.2 評価実験	11
3.2.1 実験設定	12
3.2.2 実験結果	12
3.2.3 誤り分析	13
第4章 概念語の性質に基づく順序付け	19
4.1 共起頻度に基づく概念語のスコアリング	19
4.2 評価実験	21

4.2.1	評価用データ	21
4.2.2	評価尺度	21
4.2.3	各順序付け対象における被験者間の一致度	23
4.2.4	共起関係の違いによる結果の比較	25
4.2.5	対義形容詞対を用いた結果の比較	27
4.2.6	順序付け結果の誤り分析	29
第 5 章	関連研究	33
5.1	順序付けに関する研究	33
5.2	オンデマンドな知識獲得に関する研究	34
第 6 章	おわりに	37
	発表文献	42
	付録 A	43
	付録 B	58

目 次

3.1 比較文の獲得例	11
-----------------------	----

第1章 はじめに

1.1 ウェブテキストを用いた順序付け知識獲得の重要性

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、外出中に食事をしたいと思ったとする。この時、まず候補となる料理を列挙するだろう。そして、その状況で自分が優先している観点（「安さ」や「美味しさ」など）に基づいて列挙した料理を順序付けることで最終的な意思決定を行うだろう。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。物事の順序付けという行為は人々が今までに経験した知識に基づいて無意識のうちにやっているものであるが、これを大規模ウェブテキストの統計量をもとに解く手法について検討する。ウェブテキストから物事の順序付け知識が獲得できることにより、ある特定の状況や条件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が可能となる。また、順序付け知識をアプリケーションに応用することにより、機械自身によって物事の優劣を判別できるようになるため、より人間のような柔軟な振る舞いを持つことが可能になると考えられる。加えて、ウェブテキストを解析することにより、順序付けする物事や観点の拡張も可能になる。物事の拡張は人間の思考だけでは限界があるため、ウェブテキストを用いることに意義があると考えられる。これにより、人々が物事を選択をする際に候補や観点の補完をすることができるため、人々はより良い意思決定を行うことが可能になると期待できる。また、類似する概念語を半自動で収集できるため、高度な情報収集を行うことも可能であると考えられる。

1.2 本研究の目的と貢献

本研究では、類似する複数の概念語を多様な観点で順序付けしたいという問題に対して、複数の概念語に共通する順序付けの観点となる形容詞を獲得するタスクと、複数の概念語を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の2つに分け、それを実現するための方法を提案する。

観点となる形容詞の獲得タスクにおいては、比較文を用いた形容詞の絞込みを行った後、概念語と形容詞との相互情報量を計算することで各形容詞に対して概念語集合に含まれる概念語との相互情報量の総和をスコアとすることで観点として最もらしい形容詞を獲得する。概念語の順序付けタスクを提案するにあたり、入力概念語と形容詞のペアをどのように決めるかということが課題の一つとなっていたが、観点の獲得タスクにより、以降で我々が提案する形容詞との共起に基づく概念語の順序付けタスクの入力を半自動的に獲得をすることができる。

概念語を順序付けタスクにおいては、(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といった課題が存在する。(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。

1.3 本論文の構成

本論文の構成は次のとおりである。

第2章 本研究で用いる共起関係の定義と頻度獲得の方法，また二つのタスクで共通となるデータセットの説明をする．

第3章 概念語集合から，順序付けの観点となる形容詞の獲得手法を提案し，その評価を行う．

第4章 概念語を形容詞の性質に基づいて順序付けする手法を提案し，その評価を行う．

第5章 関連研究についての説明と本研究の位置づけについて説明する．

第6章 全体のまとめと今後の課題について述べる。

第2章 アプローチ

本研究では類似する複数の概念語を多様な観点で順序付けするという問題に対して、複数の概念語に共通する順序付けの観点となる形容詞をオンデマンドに獲得するタスクと、複数の概念語を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の2つに分けることでそれを実現するための方法を本研究では提案する。具体的には、まず複数の概念語集合を評価データとして作成し、前半のタスクで概念語集合ごとに形容詞のペアを生成する。後半のタスクでは前半のタスクで獲得された概念語集合と形容詞のペアを入力として与え、概念語集合に含まれる各概念語の具体的な順序付けを行う。手法としては概念語と形容詞との共起頻度をもとに順序付けの観点となる形容詞の獲得や概念語の順序付けを行う。

本章では両タスクで共通している、共起頻度の獲得方法と評価用データの作成手順について説明する。

2.1 共起頻度獲得における表記揺れの統合

例えば、「ネコは可愛い」と「猫はかわいい」はどちらも猫の可愛さについて言及している文であるが、単純に名詞と形容詞の共起頻度としてカウントすると別のものとして扱われてしまう。本研究では共起頻度によって概念語と形容詞の結びつきを測るため、このような状況は望ましくない。そこで形容詞、概念語それぞれについて以下のようにして正規化をした。

形容詞についてはカタカナを平仮名に正規化し、フリーで配布されている Juman 辞書¹を利用することで最終的に漢字表記に統合した。

¹<https://mecab.googlecode.com/files/mecab-jumandic-7.0-20130310.tar.gz>

2.2. 概念語と形容詞の共起関係の獲得

概念語に関しては入力として与える概念語が既に決まっているため、各概念語ごとに容易に想起できる表記揺れ（「ウィスキー」と「ウイスキー」など）を辞書として作成することで最終的な概念語の表記揺れの統合を行った。

2.2 概念語と形容詞の共起関係の獲得

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2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2種類の共起頻度を予め計算し、優劣の検証が行えるようにする。

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。

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。

2.3. 順序付け対象となる概念語集合の獲得

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。主格係り受けを共起関係として利用することによりこのような不適切な文を除外することができる。

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 2.1 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。

本研究で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 月までに記述されたブログ記事 (約 2 億記事, 約 19 億文) を利用した。解析器として、形態素解析には Kaji らの手法 [5], 係り受け解析には Yoshinaga らが開発した J.DepP[6][7] を利用した。また、共起頻度の集計対象となる文の質を確保するため、特殊記号「?」、副詞「案外、意外、まるで」、その他「のかな、だろうか、なのに、のような、みたいな」を含む文は共起関係をカウントする対象から除外した。

2.3 順序付け対象となる概念語集合の獲得

ある概念語を入力として与え、比較可能な類似する概念語を Brown クラスタリングによって獲得し、順序付け対象となる概念語集合を作成する。Brown クラスタリングは階層クラスタリングの一種であり、根ノードからのパスによってクラスタが分類されている。本研究では比較可能な類似する概念語は同一クラスタ内に属する概念語同士と定義をする。概念語集合は以下のように獲得をする。なお、Brown ク

2.3. 順序付け対象となる概念語集合の獲得

文	共起関係	
	同一文内共起	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 2.1: 共起頻度の集計対象となる文

ラストリングを作成するためのテキストデータには2009年のブログ記事からランダムに1/10抽出したテキストを使用している。

まず入力として与えた概念語の属するクラスタのパスを求め、次に、同一クラスタ内に属する単語を高頻度順に並べる。最後に、入力として与えた概念語が属するカテゴリと同族の単語を同一クラスタから選ぶ。例えば、入力として「クジラ」を与えた場合、「動物」というカテゴリが想起できるので、動物名を同一クラスタから選択する。以上の手続きによって最終的に作成した評価データを表 2.2 にまとめる。以降の説明ではこの「カテゴリ」名を概念語の集合名として説明をする。

次節以降で使用するデータセットは表 2.2 を共通して使用する。はじめのタスクで各カテゴリを順序付けする観点として適切な形容詞を獲得し、次のタスクで前のタスクで獲得したカテゴリと形容詞のペアを入力として与えることで共起頻度に基づく概念語の順序付けを行う。

2.3. 順序付け対象となる概念語集合の獲得

カテゴリ	概念語集合
動物	ネコ, ライオン, クマ, ウシ, ネズミ, クジラ, キリン, イヌ, ゾウ, サル
乗り物	飛行機, 電車, バス, 船, 新幹線, タクシー, 自転車, ヘリコプター 自動車
酒	ビール, ワイン, 焼酎, カクテル, シャンパン, ウイスキー, ハイボール チューハイ, マッコリ
食べ物	カレー, 焼きそば, ハンバーガー, チャーハン, パン, 寿司, ピザ, ステーキ パスタ, ラーメン
スポーツ	野球, ゴルフ, サッカー, テニス, プロレス, 相撲, バレー, 水泳, ラグビー ボクシング
果物	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
家具	ストーブ, エアコン, ミシン, 扇風機, ドライヤー, ヒーター
職業	パティシエ, アーティスト, ミュージシャン, 力士, パイロット, ピアニスト エンジニア, ボクサー, セラピスト, ジョッキー, レーサー
国	アメリカ, ロシア, 中国, イギリス, フランス, ブラジル, インド, カナダ イタリア, ドイツ
都市	大阪, 京都, 横浜, 名古屋, 沖縄, 福岡, 札幌, 神戸, 仙台
貴金属	ネックレス, 指輪, イヤリング, ピアス, ブレスレット
コンピュータ	パソコン, カメラ, プリンター, スマートフォン, カーナビ, ビデオカメラ
ゲーム機	D S, W i i, P S P, ファミコン
服	スーツ, 着物, 浴衣, シャツ, ジャージ, パーカー, セーター, カットソー
企業	アマゾン, ヤフー, グーグル
楽器	ギター, ピアノ, バイオリン, フルート, チェロ, トランペット, オルガン ウクレレ
花	サクラ, ウメ, モミジ, チューリップ, アジサイ, コスモス, イチョウ ヒマワリ, タンポポ
消耗品	シャンプー, トイレットペーパー, ティッシュ, ストロー
運動	ダイエット, 筋トレ, ジョギング, ウォーキング, エクササイズ
賭博	パチンコ, 競馬, 競艇, パチスロ
嗜好品	タバコ, コーヒー, 茶, チョコレート, コーラ, ビール, ガム
お菓子	チョコレート, アイスクリューム, ホットケーキ, ポップコーン ポテトチップス, アップルパイ
記念日	バレンタインデー, ホワイトデー, クリスマス, ハロウィン, バースデー イースター
建物	学校, ホテル, スーパー, マンション, コンビニ, カフェ, 幼稚園, レストラン ジム
科目	数学, 歴史, 古典, 地理, 音楽, 政治
野球チーム	ソフトバンク, カープ, ライオンズ, ベイスターズ

表 2.2: カテゴリ別の概念語集合

第3章 順序付け観点の形容詞獲得

与えられた複数の概念語を順序付けする際に必要となる観点の獲得を自動で獲得する手法について提案する。例えば、入力として「クジラ」、「ゾウ」、「キリン」のようなあるカテゴリをすることのできる複数の概念語を入力として与え、それらを順序付けする観点となる形容詞（「大きい」や「重い」や「可愛い」など）を出力する。なお、概念語は、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）とする。

3.1 提案手法

我々は、概念語集合を順序付けする観点となる形容詞を概念語と形容詞との共起頻度に基づいて獲得する手法を提案する。概念語集合との共起頻度をもとに形容詞の結びつきの強さをスコア化することで最終的な出力を獲得する。共起頻度によって順序付けの観点を獲得する理由としては、(1) 複数の概念語間で共通している性質は各概念語に対して多く共起をしやすく、(2) 類似する概念語同士はそれぞれ似たような形容詞と多く共起をしやすい、という二つの直感に基づくものである。

しかしながら、概念語集合との結びつきの強い形容詞が順序付けの観点として適切かどうかは限らない。そこで提案手法では、順序付けの観点となる形容詞は物事を比較する際に用いられるはずであるという仮定のもと、比較文を用いたフィルタリングを行い、その後、共起頻度に基づくスコアリングをすることで、信頼度の高い出力を獲得することを目指す。以下、比較文を用いた形容詞の選定方法と、概念語と形容詞との共起頻度に基づいたスコアリングの方法について説明する。

3.1.1 比較文を用いたフィルタリング

順序付けの観点となる形容詞は比較文中で表現されるという直感のもと，比較文で使われている形容詞と概念語の獲得を行った．本研究での比較文の定義は以下のようになっている

- (1) 1文中で，ある形容詞に対して2つの概念語がそれぞれ係り受け関係にある．
- (2) 片方の概念語が形容詞と主格係り受け関係，もしくは「方（ほう）が」を挟んで形容詞と係り受け関係にある
- (3) もう片方の概念語の同一文節内に助詞「より」を含んでいる，もしくは「比較」，「比べる」を挟んで形容詞と係り受け関係にある

具体的な例を図 3.1(a) と 3.1(b) に示す．

3.1.2 共起頻度に基づくスコアリング

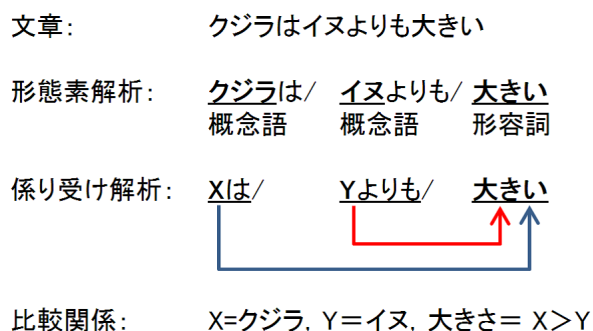
入力された概念語を順序付けする際に適した形容詞であるかを評価するために，概念語との共起頻度をもとに形容詞をスコア化する．計算されたスコアが高い形容詞ほどより順序付けの観点として適したものとする．単純に概念語と形容詞との共起頻度だけを求めると「すごい」のような高頻度な形容詞が上位にくることになり，入力で与えた概念語集合を順序付けする際の特徴的な性質を記述した形容詞を獲得することは難しい．

そこで，スコア付けには単語同士の結びつきの強さを示す自己相互情報量 (PMI) を用いることで，形容詞の出現頻度を考慮したスコア付けを行う．

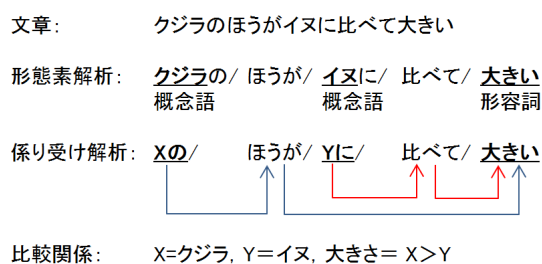
$$\text{Score}(\text{形容詞}) = \frac{1}{|N|} \sum_{n \in N} \text{PMI}(\text{概念語 } n, \text{形容詞}) \quad (3.1)$$

$$\text{PMI}(\text{概念語}, \text{形容詞}) = \log_2 \frac{P(\text{概念語}, \text{形容詞})}{P(\text{形容詞})P(\text{概念語})} \quad (3.2)$$

3.2. 評価実験



(a) 形容詞との直接の係り受け関係がある場合



(b) 形容詞と間接的に係り受け関係がある場合

図 3.1: 比較文の獲得例

ここで、 N は 3.1 節で拡張することで得られた概念語集合を表しており、 $|N|$ はその要素数、 $P(\cdot)$ は単語の出現確率を示す。このようにして得られたスコアの高い形容詞を出力とする。なお、共起頻度をカウントする際の共起関係として同一文内共起を用いる。

3.2 評価実験

本章では、前章で述べた提案手法を用いて概念語集合からそれらを順序付け可能な形容詞の獲得を行う。スコアの上位 10 件を示すことで提案手法の有用性について検証する。獲得する対象となる形容詞は頻度上位 1,000 件の形容詞とした。出力対

3.2. 評価実験

象の形容詞を頻度によって選定する理由としては、観点となり得る形容詞は（例えば「大きい」など）は、テキストデータ上で高頻度であるはずという直感に基づくものである。

また、適合率を正しく評価するために、形容詞と反義形容詞は同じ観点であるとみなし、対義形容詞対（形容詞と反義形容詞のペア）の頻度は頻度が多い形容詞に統一をした。

3.2.1 実験設定

提案手法によって獲得された形容詞が順序付けの観点として適切であるかは4人の被験者に評価をしてもらうことで適合率を求めた。被験者に各カテゴリごとに獲得された形容詞10件の正誤判定を行ってもらった。その結果をもとに最上位の適合率 (precision@1), 上位5件までの適合率 (precision@5), 全ての出力に対する適合率 (precision@10) の4人の平均を評価値とした。

3.2.2 実験結果

本研究では比較表現を利用せず、共起頻度によるスコアリングのみを用いた手法をベースライン、比較表現を利用し形容詞を選定した後に共起頻度によるスコアリングを行う手法を提案手法として比較をした。なお、今回の実験では共起関係として同一文内共起を利用した。表3.1に各カテゴリごとに提案手法とベースラインで獲得された形容詞の precision@n の評価値を示す。表3.1は提案手法の precision@10 の値をもとに降順によって並べた。各カテゴリの適合率の平均をみると precision@n は全ての場合において提案手法のほうがベースラインより優っているという結果になった。このことから、比較表現によって形容詞を選定することが順序付け観点の獲得においては有用であることがわかった。

表3.3には各カテゴリごとに獲得された形容詞の具体例を示す。この一覧も表3.1と同様に提案手法における precision@10 の値を降順に並べる。カテゴリに着目を見ると、食べ物や都市や乗り物など日常生活において比較する機会の多いものが上位

3.2. 評価実験

に來ていることがわかる．このことから，比較する際に用いられる形容詞は順序付けの観点として適切であるということがわかる．しかしながら，具体的な形容詞についてしてみると，食べ物に関しては「美味しい」「おいしい」，都市に関しては「暖かい」「暖かだ」といった同義語が多く含まれており，観点の種類に重複が見られることがわかる．次節では誤り分析について述べる．

3.2.3 誤り分析

カテゴリ自体の不適合さ

今回は Brown クラスタからあるカテゴリを想起できるものをランダムにサンプリングした．その中で適合率の低い結果となったカテゴリは表 3.1 より消耗品や建物や家具などが挙げられる．これらは適合率が上位の乗り物や食べ物とは違い，概念語同士を比較，順序付けする機会がそもそも少ない．その結果，それらのカテゴリとの結び付きが強い形容詞が順序付けをする際の観点としては不適切なものであったということが考えられる．

解析誤り

表 3.3 に着目をするすると，ナ形容詞の解析誤りが多く見受けられた．例えば，ゲーム機における「ソフトだ」，お菓子における「買い込んだ」，都市における「行方不明だ」がそれにあたる．これら解析誤りの単語についてはカテゴリにおける特徴的な単語であるため，共起頻度を用いた結び付きの強さのスコアリング方法を改善するだけでは除去が難しいと考えられる．そのため，比較表現のような直感的なフィルタリング等をあらたに手法として加える必要がある．さらに，解析誤りにおいては以下のような文章も比較文として判定されてしまい，フィルタリングを通過したと考えらるので比較対象の品詞についても考慮をする必要があると考えられる．

- 今日大阪より お世話になっていた方からプレゼントをもらった
- カーナビはハードよりもソフトだ

3.2. 評価実験

- 何よりもビールはたんと買い込んでおきましたのでひもじい思いはしなくて済みました.

観点の不明瞭さ

表 3.3 における, 科目の「顕著だ」や国の「深刻だ」などがそれにあたる. 本来ならば主格を伴って用いられる形容詞を単体で順序付けの観点として利用することは難しいと考えられる. この解決としてはフレーズ（「経済が深刻だ」など）を観点として獲得することが有効だと考えられる.

概念語特有の観点の抽出

表 3.3 におけるカテゴリ花の「黄色い」やカテゴリ家具の「うるさい」がそれにあたる. カテゴリ内に含まれる全ての概念語との相互情報量は高くなくとも一つ（もしくは数個）の概念語との相互情報量のスコアが高いと獲得されてしまうのが原因と考えられる. 単純なスコアの総和でなく, 相互情報量を用いているのである程度は緩和されていると考えられるがそれでも改善の必要があることがわかる. このためにはスコアリングの手法についてもう少し検討をする必要がある.

3.2. 評価実験

カテゴリ	<i>precision@n</i>					
	<i>n</i> = 1		<i>n</i> = 5		<i>n</i> = 10	
	ベースライン	提案手法	ベースライン	提案手法	ベースライン	提案手法
食べ物	100.0	100.0	60.0	80.0	52.5	70.0
都市	100.0	100.0	60.0	85.0	55.0	67.5
乗り物	0.0	100.0	20.0	65.0	22.5	65.0
ゲーム機	0.0	0.0	35.0	60.0	40.0	65.0
果物	75.0	50.0	70.0	70.0	67.5	60.0
野球チーム	0.0	100.0	70.0	75.0	52.5	60.0
スポーツ	0.0	0.0	35.0	55.0	50.0	57.5
記念日	0.0	75.0	60.0	80.0	55.0	57.5
科目	0.0	25.0	15.0	60.0	32.5	57.5
賭博	0.0	25.0	25.0	40.0	30.0	55.0
企業	100.0	100.0	50.0	40.0	37.5	52.5
お菓子	0.0	25.0	50.0	65.0	50.0	47.5
酒	100.0	100.0	60.0	75.0	65.0	45.0
動物	75.0	75.0	55.0	70.0	47.5	42.5
貴金属	75.0	50.0	45.0	50.0	50.0	42.5
花	0.0	0.0	25.0	60.0	30.0	37.5
コンピュータ	50.0	50.0	50.0	55.0	50.0	35.0
服	50.0	75.0	65.0	35.0	42.5	35.0
運動	50.0	0.0	35.0	25.0	30.0	35.0
国	50.0	0.0	35.0	40.0	40.0	32.5
楽器	75.0	0.0	65.0	25.0	47.5	32.5
嗜好品	25.0	25.0	35.0	20.0	30.0	27.5
家具	0.0	0.0	10.0	20.0	17.5	22.5
職業	0.0	0.0	5.0	5.0	17.5	20.0
建物	0.0	0.0	5.0	10.0	7.5	17.5
消耗品	0.0	0.0	5.0	20.0	2.5	10.0
平均	35.6	41.3	40.2	49.4	39.3	44.2

表 3.1: *precision@n* による獲得された形容詞の評価
(全員正例)

3.2. 評価実験

カテゴリ	<i>precision@n</i>					
	<i>n</i> = 1		<i>n</i> = 5		<i>n</i> = 10	
	ベースライン	提案手法	ベースライン	提案手法	ベースライン	提案手法
食べ物	100.0	100.0	60.0	100.0	60.0	70.0
都市	0.0	100.0	20.0	80.0	20.0	80.0
乗り物	100.0	100.0	100.0	100.0	90.0	70.0
ゲーム機	100.0	100.0	60.0	80.0	70.0	80.0
果物	0.0	0.0	40.0	80.0	60.0	80.0
野球チーム	100.0	100.0	100.0	100.0	100.0	80.0
スポーツ	0.0	0.0	20.0	40.0	30.0	40.0
記念日	0.0	0.0	20.0	0.0	40.0	30.0
科目	100.0	0.0	40.0	40.0	70.0	30.0
賭博	100.0	100.0	80.0	100.0	70.0	80.0
企業	100.0	100.0	60.0	80.0	80.0	70.0
お菓子	100.0	100.0	80.0	80.0	70.0	50.0
酒	0.0	0.0	40.0	80.0	50.0	80.0
動物	100.0	100.0	80.0	80.0	70.0	90.0
貴金属	100.0	100.0	60.0	60.0	50.0	70.0
花	100.0	0.0	100.0	60.0	90.0	70.0
コンピュータ	0.0	0.0	60.0	80.0	60.0	50.0
服	0.0	0.0	20.0	40.0	10.0	20.0
運動	100.0	0.0	80.0	60.0	80.0	60.0
国	0.0	100.0	60.0	80.0	60.0	80.0
楽器	100.0	100.0	80.0	40.0	60.0	50.0
嗜好品	0.0	100.0	80.0	80.0	70.0	60.0
家具	0.0	100.0	80.0	80.0	90.0	50.0
職業	0.0	0.0	20.0	20.0	30.0	30.0
建物	0.0	0.0	40.0	60.0	60.0	50.0
消耗品	0.0	100.0	80.0	100.0	60.0	80.0
平均	50.0	57.7	60.0	69.2	61.5	61.5

表 3.2: *precision@n* による獲得された形容詞の評価
(一人以上正例)

3.2. 評価実験

表 3.3: 獲得された具体的な形容詞

カテゴリ	形容詞リスト
食べ物	美味しい, 美味しい, 美味だ, 満腹だ, 旨い, ヘルシーだ, ジューシーだ, たっぷりだ, リーズナブルだ, 高級だ
都市	蒸し暑い, 肌寒い, 寒い, 暖かい, 恋しい, 有名だ, 近い, 暖かだ, 世話だ, 行方不明だ
乗り物	快適だ, 高速だ, 間近だ, 大形だ, うるさい, 便利だ, 危ない, 速い, 近い, 騒がしい
ゲーム機	ソフトだ, 夢中だ, 欲しい, 手軽だ, しょぼい, おもしろい, めんどい, 面白い, 格段だ, 新しい
果物	甘酸っぱい, ジューシーだ, みずみずしい, 酸っぱい, フレッシュだ, さわやかだ, 美味しい, 爽やかだ, 小振りだ, 辛い
野球チーム	好調だ, 強い, 有利だ, 苦しい, 熱い, 予想外だ, 厳しい, 楽しみだ, 明らかだ, 酷い
スポーツ	夢中だ, メジャーだ, 盛んだ, 好きだ, 熱い, 面白い, おもしろい, 熱心だ, 大好きだ, 大嫌いだ
記念日	特別だ, 待ち遠しい, 盛大だ, スペシャルだ, 素敵だ, 寂しい, 可愛い, 近い, めでたい, いいだ
科目	無知だ, 偉大だ, 苦手だ, 興味深い, 難解だ, 身近だ, 緻密だ, 重要だ, 顕著だ, 深い
賭博	詳しい, 健全だ, 熱い, 堅実だ, 有利だ, 好きだ, 面白い, 簡単だ, 忙しい, おもしろい
企業	便利だ, 詳しい, 可能だ, 正確だ, 格安だ, 詳細だ, 高い, 簡単だ, 巨大だ, 見易い
お菓子	香ばしい, 美味しい, 美味だ, 買い込んだ, 美味しい, ヘルシーだ, 辛い, 大好きだ, たっぷりだ, 控え目だ
酒	美味しい, 美味しい, 旨い, 美味だ, 苦い, リーズナブルだ, ぴったりだ, リッチだ, 豊富だ, 辛い
動物	愛らしい, 賢い, でっかい, 可哀相だ, 可愛い, 大喜びだ, 巨大だ, 大形だ, そっくりだ, 器用だ
貴金属	可愛い, 欲しい, 高価だ, 細い, いいだ, クールだ, 地味だ, 大切だ, 美しい, 大事だ
花	黄色い, 鮮やかだ, 可憐だ, 綺麗だ, 淡い, 見事だ, 青い, 真っ赤だ,

3.2. 評価実験

表 3.3: 獲得された具体的な形容詞

カテゴリ	形容詞リスト
	眩しい, 華やかだ
コンピュータ	便利だ, 大形だ, 高価だ, 手軽だ, 新しい, 好調だ, 欲しい, ソフトだ, 高速だ, 見易い
服	薄手だ, タイトだ, 洒落だ, 寒い, 赤い, シンプルだ, 色っぽい, 地味だ, 欲しいぴったりだ
運動	健康だ, ハードだ, 手軽だ, 最適だ, きつい, 軽い, しんどい, 気持ち良い, 有効だ, 楽だ
国	盛んだ, 優位だ, 根強い, 厳格だ, 有名だ, 強固だ, 深刻だ, 貧乏だ, 著しい, 興味深い
楽器	多彩だ, 重厚だ, 力強い, 素晴らしい, 上手だ, 美しい, 巧い, 素敵だ, シンプルだ, ステキだ
嗜好品	苦い, ほのかだ, 買い込んだ, 美味しい, 格別だ, 旨い, 美味しい, マイルドだ, 辛い, 冷たい
家具	寒い, 暖かい, 快適だ, うるさい, 冷たい, 苦手だ, 邪魔だ, 臭い, 気持ち良い, 欲しい
職業	優秀だ, 若い, 素晴らしい, 最高だ, 未熟だ, 孤独だ, 好きだ, 過酷だ, 真剣だ, 抜群だ
建物	高級だ, 買い込んだ, おしゃれだ, 便利だ, 普通だ, 格安だ, 騒がしい, 近い, 一緒だ, 親切だ
消耗品	細い, 上手だ, 柔らかい, 痒い, 軽い, 便利だ, 優しい, 楽だ, 大人しい, 薄い

第4章 概念語の性質に基づく順序付け

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。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、入力の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。

4.1 共起頻度に基づく概念語のスコアリング

前節では、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ことを説明した。しかしながら、各概念語の出現頻度は当然ながら異なるため、単純に形容詞との共起頻度のみでは正しく概念語の性質の強さを比較することができない。よって、以下のように各概念語自体の出現頻度を考慮することで、概念語と形容詞との共起頻度を正規化を行い、同じ尺度で順序付けをする必要がある。

$$\text{順序付けのスコア} = \frac{\text{概念語と形容詞の共起頻度}}{\text{概念語の出現頻度}} \quad (4.1)$$

4.1. 共起頻度に基づく概念語のスコアリング

提案手法では式 4.1 の概念を単語同士の結び付きの強さを表す相互情報量を用いたスコアリング手法によって実現する．ある 2 つの単語， $word_1$ と $word_2$ の相互情報量は以下のように計算できる．

$$PMI(word_1 \& word_2) = \log_2 \frac{p(word_1 \& word_2)}{p(word_1) p(word_2)} \quad (4.2)$$

ここで $p(word_1)$ は文書における単語 $word_1$ の出現確率， $p(word_1 \& word_2)$ は文書における $word_1$ と $word_2$ の共起確率を表している．

式 (4.2) の $word_1$ を「カレー」， $word_2$ を「安い」とすることで，カレーの安さの程度をスコア化できる．この場合，形容詞（安い）との相互情報量をもとにした概念語の順序付けが可能となる．

本研究では，形容詞の相互情報量に加え，反義形容詞（「安い」の場合の「高い」）との相互情報量も考慮した，対義形容詞対（「安い」と「高い」の組み合わせ）との相互情報量を用いた概念語の順序付け手法を提案する．スコアには概念語と対義形容詞対との相互情報量をそれぞれ計算し，その差を用いる．対義形容詞対との相互情報量を考慮することにより，単一形容詞では評価ができなかった，一般的なカレーの値段は安い部類に位置するのか，もしくは，高い部類に位置するのかといった事象がスコア化でき，より適切な順序付けが可能になると考えられる．対義形容詞対を用いたスコア付けは以下のように計算される．

$$\text{概要語と形容詞との } PMI - \text{概念語と反義形容詞との } PMI \quad (4.3)$$

式 (4.2)，(4.3) より最終的にスコアは以下のように計算される．

$$\log_2 \frac{\text{概念語と形容詞の共起頻度} \times \text{反義形容詞の出現頻度}}{\text{概念語と反義形容詞の共起頻度} \times \text{形容詞の出現頻度}} \quad (4.4)$$

共起頻度に関しては，ゼロ頻度問題を回避するために共起頻度に 1 を加算し，スムージングを行う．本研究における反義形容詞は，表記揺れを解消するときに使用した

4.2. 評価実験

Juman 辞書に定義されている単語を使用した．反義形容詞が定義がされていない形容詞は反義語が存在しないということにした．得られたスコアをもとに降順で概念語を整理し順序付けを行う．

4.2 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，順序付けに用いる手がかりの優劣を検証する．本研究では，(1) 共起関係の違いによる手法の優劣，(2) 対義形容詞対による手法の優劣，の二つの点において実験結果より検証する．

4.2.1 評価用データ

実験に先立ち，評価用のデータを作成した．実際に順序付けする対象となる概念語集合は表 2.2 に掲載している概念語集合を使用する．順序付けをする観点となる形容詞は前章のタスクで全ての被験者が順序付け可能な観点と評価した観点を利用する．このようにして各カテゴリに対して順序付け観点のペアを作成した．カテゴリと観点のペアを表 4.1 に示す．また，括弧内に記入されている形容詞は反義形容詞である．これらのペアを評価用データとし，順序付けをした．さらに反義形容詞が定義されている観点に関しては対義形容詞対による順序付けの評価用データとして利用した．

4.2.2 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 つの順列を X と Y ， X と Y における概念語 i の順位差を D_i ， X と Y において同順位の概念語の個数

4.2. 評価実験

カテゴリ	形容詞リスト
動物	可愛い
乗り物	快適だ, 速い (遅い)
酒	美味しい
食べ物	美味しい, リーズナブルだ
スポーツ	メジャーだ (マイナーだ)
運動	面白い (つまらない)
果物	酸っぱい, 美味しい
国	有名だ (無名だ ¹), 貧乏だ (裕福だ)
都市	寒い (暑い)
貴金属	高価だ (安価だ)
コンピュータ	高価だ (安価だ)
ゲーム機	しょぼい, 面白い (つまらない), 新しい (古い)
企業	便利だ (不便だ), 簡単だ (複雑だ), 巨大だ
花	綺麗だ
賭博	好きだ, 面白い (つまらない), 簡単だ (複雑だ)
お菓子	美味しい, 大好きだ
記念日	素敵だ
科目	苦手だ (得意だ), 重要だ
野球チーム	好調だ (不調だ), 強い (弱い), 酷い

表 4.1: カテゴリと形容詞の組み合わせ

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.

$$\rho = \frac{T_x + T_y - \sum_{i=1}^N D_i^2}{2\sqrt{T_x T_y}} \quad (4.5)$$

$$T_x = \frac{N^3 - N - \sum_{i=1}^{n_x} (t_i^3 - t_i)}{12} \quad (4.6)$$

$$T_y = \frac{N^3 - N - \sum_{j=1}^{n_y} (t_j^3 - t_j)}{12} \quad (4.7)$$

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

4.2. 評価実験

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．

4.2.3 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．表 4.2 に，被験者の順位相関係数の平均を示す．なお，相関係数の上位5位と下位5位については太字でその値を示す．

相関係数の高い順序付け結果となった組み合わせに見られる傾向は二つあると考えられる．まず，一つ目は順序付け対象となる概念語の数が少ないカテゴリが上位に来やすいということである．例えば，企業の便利さの順序付け結果は全ての被験者で結果が同じ（相関係数 1.000）であったり野球チームの強さも相関係数が高い結果となっている．これは，カテゴリ「企業」には3つしか概念語，「野球チーム」には4つしか順序付け対象が存在していないため，順序付けの一致する可能性が高くなると言える．二つ目は定量的に評価できる観点であるということである．例えば，都市や乗り物の概念語の数は多いが，定量的に評価できる気温や速度に基づく順序付けであるため，相関係数が高くなっている．

一方，相関係数が低い順序付けの結果となった組み合わせに見られる傾向としては，相関係数が高い順序付けとは逆に主観に基づく順序付けであるということである．簡単さや面白さといった人によって基準が異なる観点をもとに順序付けした結果，相関係数が低くなったと考えられる．また，貴金属のような概念語によって程度の強さに大きな幅がある（例えば，指輪とひと言でいっても高いものから低いものまで多く存在する）ものを順序付けすると値段という定量的な観点にも関わらず，相関係数は低くなることがわかる．

4.2. 評価実験

カテゴリ	形容詞	ρ (平均)
動物	可愛い	0.405
乗り物	快適だ	0.542
	速い	0.856
酒	美味しい	0.095
食べ物	美味しい	0.399
	リーズナブルだ	0.577
スポーツ	メジャーだ	0.748
	面白い	0.48
果物	酸っぱい	0.649
	美味しい	0.409
国	有名だ	0.591
	貧乏だ	0.105
都市	寒い	0.917
貴金属	高価だ	-0.174
コンピュータ	高価だ	0.696
	しょぼい	-0.162
	面白い	-0.233
ゲーム機	新しい	0.894
企業	便利だ	1.000
	簡単だ	0.000
	巨大だ	0.667
花	綺麗だ	0.571
賭博	好きだ	0.100
	面白い	0.100
	簡単だ	-0.146
お菓子	美味しい	0.574
	大好きだ	0.612
記念日	素敵だ	0.441
科目	苦手だ	0.591
	重要だ	0.337
野球チーム	好調だ	0.757
	強い	0.811
	酷い	0.400

表 4.2: 被験者の付けた順位の平均順位相関係数

4.2.4 共起関係の違いによる結果の比較

共起関係として同一文内共起を利用した場合と、主格係り受けを利用した場合の二通りを比較し、共起関係による優劣を検証する。表 4.3 に被験者によってつけられた順序と提案手法によって順序付けされた順序と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均値の上限値も併せて示す。表 4.3 から分かるように係り受けを共起関係として利用したほうが同一文内共起よりも相関係数の平均が高いことがわかる。しかしながら、その差はわずかなものであり、有用性が確認できたとはいうのは難しい。これは、前述したように、共起関係は質・量のトレードオフの性質があるため、有意な差が観測できなかったのではないかと推測できる。また、形容詞に着目すると美味しいといった主観に基づく順序付けの相関は低いことがわかる。(表 6.4, 表 6.5)。

なお、本研究で行った同一文内共起と係り受け関係との具体的な順序付け結果の比較は付録 A に示す。

4.2. 評価実験

カテゴリ	形容詞	上限値	共起関係	
			文内共起	係り受け
動物	可愛い	0.654	0.170	0.050
乗り物	快適だ	0.777	0.355	0.385
	速い	0.917	0.090	0.356
酒	美味しい	0.528	-0.173	-0.144
食べ物	美味しい	0.733	-0.001	-0.108
	リーズナブルだ	0.796	-0.757	-0.651
スポーツ	メジャーだ	0.888	-0.230	-0.371
	面白い	0.773	-0.232	-0.077
果物	酸っぱい	0.828	0.786	0.763
	美味しい	0.741	-0.113	0.278
国	有名だ	0.818	0.220	-0.141
	貧乏だ	0.585	0.073	0.055
都市	寒い	0.917	0.251	0.375
貴金属	高価だ	0.478	0.336	-0.311
ゲーム機	高価だ	0.825	-0.276	0.042
	しょぼい	0.355	0.250	0.355
	面白い	0.250	0.050	0.200
	新しい	0.948	0.743	-0.384
企業	便利だ	1.000	0.500	0.500
	簡単だ	0.500	0.250	0.250
	巨大だ	0.750	0.750	0.750
花	綺麗だ	0.806	0.350	0.517
賭博	好きだ	0.550	-0.350	-0.350
	面白い	0.550	-0.350	0.250
	簡単だ	0.495	-0.332	-0.068
お菓子	美味しい	0.793	0.206	-0.139
	大好きだ	0.807	-0.139	0.105
記念日	素敵だ	0.702	0.294	0.307
科目	苦手だ	0.835	-0.900	-0.900
	重要だ	0.712	0.213	-0.236
野球チーム	好調だ	0.898	0.387	0.176
	強い	0.924	-0.395	0.369
	酷い	0.724	0.313	0.571
平均		0.723	0.071	0.084

表 4.3: 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

4.2.5 対義形容詞対を用いた結果の比較

反義形容詞が定義されている形容詞に着目し，単一形容詞によるスコアリングの場合と，対義形容詞対を用いたスコアリングの場合の二通りを比較し，反義形容詞を利用する有用性について検証する．表 4.4 に被験者によってつけられた順序と提案手法によって順序付けされた順序と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均値の上限値も併せて示す．なお，表中の手法について，「形容詞」は式 4.1 を用いて単一形容詞との相互情報量をもとにスコア付けに利用した手法，「+ 反義」は式 4.4 を用いて対義形容詞対の相互情報量をもとにスコア付けに利用した手法である．

表 4.4 をみると，提案手法である「+ 反義」のほうがベースラインである「形容詞」よりも相関係数の平均が高いことがわかる．実験の結果，反義形容詞を用いることの有用性が確認できた．

なお，本研究で行った対義形容詞対を用いた手法とベースラインとの具体的な順序付け結果の比較は付録 B に示す．

次節では，実験結果の中で相関係数に極端な差が出た部分や結果が悪い部分に着目し，どのような概念語と形容詞の共起文が結果に影響を与えているのかを分析し，考察をする．

4.2. 評価実験

カテゴリ	対義形容詞対		上限値	手法	
				形容詞	+ 反義
乗り物	速い	遅い	0.917	0.090	0.082
スポーツ	メジャーだ	マイナーだ	0.888	-0.230	0.044
	面白い	つまらない	0.773	-0.232	-0.058
国	有名だ	無名だ	0.818	0.220	0.763
	貧乏だ	裕福だ	0.585	0.073	0.204
都市	寒い	暑い	0.917	0.251	0.311
貴金属	高価だ	安価だ	0.478	0.336	0.429
コンピュータ	高価だ	安価だ	0.825	-0.276	0.087
ゲーム機	面白い	つまらない	0.250	0.050	0.200
	新しい	古い	0.948	0.743	0.948
企業	便利だ	不便だ	1.000	0.500	1.000
	簡単だ	複雑だ	0.500	0.250	-0.250
賭博	好きだ	嫌いだ	0.550	-0.350	0.250
	面白い	つまらない	0.550	-0.350	0.350
	簡単だ	複雑だ	0.495	-0.332	0.395
科目	苦手だ	得意だ	0.835	-0.900	-0.153
野球チーム	好調だ	不調だ	0.898	0.387	0.687
	強い	弱い	0.924	-0.395	0.724
平均			0.725	-0.023	0.309

表 4.4: 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

4.2.6 順序付け結果の誤り分析

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。本章では誤りについて、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。

- 先日いった**沖縄**は想像以上に**寒**かった。
- **パイナップル**がが思いのほか**甘**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、沖縄が寒い、パイナップルは甘い果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。

- 夏の**北海道**はすごい**暑**いんだって。
- 今の**自転車**は**速**い。
- ここは、**寿司**がとても**リーズナブル**だ。
- アメリカだと**サッカー**は**マイナー**なスポーツとのことだ。

例えば最初の文は、「本来の北海道は寒いですが、夏に限っては暑い」という意味であり、北海道の(普遍的な)気温を記述した文ではない。これらの文脈上でのみ成立する概念語の性質を共起頻度としてカウントしているため、順序付けが不自然な結果になったと考えられる。これについては概念語の修飾語を考慮する必要がある。

形容詞の多義性

今回の実験対象となる形容詞には存在しなかったが、形容詞の多義性による誤

4.2. 評価実験

りもあると考えられる。これは、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共起するというものである。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。
- **飛行機**が**遅い**から晩御飯は東京で食べていこう
- いつもより1本**電車**が**遅かった**ので、乗り換えのときに小走りになるんだよねー

椅子の例では、これらの「重い」は重量ではなく雰囲気のことを言っており、飛行機や電車の例では、「遅い」は速度に関するのではなく、時間に関することである。このような表面的には概念語との共起関係が成立している共起が多く観測され、これらを頻度情報としてカウントした結果、椅子は重い家具、という不自然な結果が得られてしまうと考えられる。

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「美味しさ」の他に「安さ」や「辛さ」などの性質も話題に上がりやすく、相互情報量によるスコアリングの場合、スコアが低くなり、スコアが世間一般での認識と一致しない可能性がある。

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

4.2. 評価実験

名の「モモ」などに対する文も含まれていた。今回は概念語に関してもカタカナ平仮名漢字の統合をしているため、多義性による影響は少なくなっているものの、このような誤りも観測された。他にも、「アメリカ」などの国名は主格以外での出現が多い。例えば「アメリカに来てからはずっと貧乏です。」といった文章のように、貧乏なのはアメリカではなく、アメリカに住んでいる人を指している。このような概念語自体に抽象度が高いものに関しては誤った情報を共起頻度としてカウントしてしまったと考えられる。これに関しては主格係り受けなどで概念語と形容詞の共起関係を厳密にすることが有効であると考えられる。実際、表 6.12 に示すように、共起関係として主格掛かり受けを用いたほうが、アメリカという概念語の順位が補正されたことがわかる。

ゼロ代名詞

概念語の性質を表している文ではあるが、概念語と形容詞に係り受けが存在しないもの。具体例を以下に示す。

- **ネコ**に対するイメージといえば、やはり思い浮かべるのはとにかく**可愛い**姿
- **飛行機**って高いからあまり乗らないけど、やっぱり**速い**よね。

係り受け関係ではこれらの文を共起頻度としてカウントできなかったため、順序付けに誤りが生じたと考えられる。これらの問題を解決するためには省略された形容詞の主語を補う必要がある。

相対的な比較に基づく記述

ある文脈において相対的な比較に基づき概念語の性質を記述しているもの具体例を以下に示す。

- サッカーは野球と**比べる**とマイナーな競技なんですね
- **福岡**って他の九州の県よりも寒い。

このような共起関係から頻度を計算した結果、福岡は寒い、サッカーはマイナーといった不自然な結果が導き出されたと考えられる。これに関しては比較

4.2. 評価実験

表現などを用いることにより直接的な順序付の記述を手がかりにする必要がある。

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るということが考えられる。

第5章 関連研究

5.1 順序付けに関する研究

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、レビュー文などから二つのインスタンスの優劣について言及していると思われる比較情報（～は～よりもよい）というフレーズを抽出し、それらの統計量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。これにより、例えば映画の良さといったランキングがアウトプットとして実現することができる。しかし、彼らが想定する、順序付けの尺度は「良さ」だけであるため、我々が本研究で提案している多様な観点に基づく順序付けというタスクは彼らのタスクをより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 [9]。この論文ではレビューがポジティブな評価をしているのかネガティブな評価をしているのかというタスクを行っている。各レビュー内に含まれる単語のポジネガをスコアリングすることによりそれらを総合してレビューがポジティブなことと言っているのかネガティブなことを言っているのかの判定を行っている。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 [10]。程度の強さの優劣は (X but not Y) のような実験者が予め列挙したフレーズによって獲得する（例えば、X

but not Y の場合、X よりも Y の方が程度が強い、という結論になる）。統計的にフレーズに当てはまる形容詞とその優劣を獲得した後、優劣の矛盾がある部分の誤差が最小となるような程度の順序付けをするという最小化問題に落とし込むことで最終的な形容詞の優劣を求めている。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。

5.2 オンデマンドな知識獲得に関する研究

本研究のように少数のインスタンスから類似するインスタンスと、インスタンス間に共通する性質を表す形容詞を獲得し、順序付けタスクに応用した研究はこれまで存在していないが、類似するインスタンスや関係性を共通パターンによって獲得するというタスクについてはいくつか研究がなされている.. オンデマンドな知識獲得に関する研究とは少数のシードインスタンスから類似するインスタンスを獲得し、集合を獲得することを指す。この研究分野は大きく分けて集合拡張とブートストラップ法に分類される。以下ではそれらの研究について簡単に紹介する

集合拡張は入力インスタンスと類似するインスタンスを獲得することが主な目的で、Wang や Carlson らはウェブページ上の構造に着目し、入力インスタンスに共通するラッパー (wrapper) で囲われたインスタンスを獲得することで同じ意味カテゴリに属する固有名詞 (ディズニーの映画名など) の獲得を提案している [12] [13]. 集合拡張をベイズ的アプローチによって実現する Bayesian Sets[11] という手法も提案されている。これら集合拡張の研究では、本研究と同様に入力インスタンスと類似するインスタンスの獲得が目的となっているが、本研究ではその共通するパターン (形容詞) も獲得の対象としているので、彼らの研究とは異なっていると言える。

ブートストラップ法は Yarowsky によって提案された手法であり、PMI などを用いることで入力インスタンスとの結びつきの強いパターンを抽出することで類似インスタンスを獲得をすることができる。手続きとしては主に、共通パターンの抽出、スコアリング、シードインスタンスのマージ、という過程をブートストラップ法においては行う。この一連の操作を繰り返すことで少数のシードインスタンスから類

5.2. オンデマンドな知識獲得に関する研究

似する多くのインスタンスを獲得することができる。ブートストラップ法においては入力シードや抽出パターンに制約はないため、シードインスタンスとパターンの組み合わせにより、様々な異なるタスクを行うことができ、現在までにその手法が提案されている。Yarowsky や Whiteney らは入力インスタンスが共起するフレーズをパターンとして獲得することで、語義曖昧性解消のタスクを行っている [14] [15]。また、Collins らはブートストラップ法を用いた類義語獲得を行っている [16]。

フレーズのパターンを利用した類義語獲得のタスクにおいては、あるパターンと多く共起するインスタンス同士を結びつけることでパラフレーズを獲得するという手法も提案されている [17]。さらに、インスタンスの時系列ごとの出現頻度を調べることでバーストを検出し、時系列ごとに変わるインスタンスの名称 (mp3, ウォークマン, など) をパラフレーズとして獲得するタスクも提案されている [18]。これらの手法では共起パターンとインスタンスからグラフを構築し、共起頻度をエッジの重さとしたランダムウォークによって最終的なクラスタを獲得する。ブートストラップ法では共起パターンによるグラフ構築を行うのではなく、繰り返し操作によってインスタンスを少しずつ獲得していくという手法になっている。さらにブートストラップ法では集合拡張とは異なり、入力をインスタンスのペアにすることで二つのインスタンスが共起するフレーズをパターンとして抽出できるため、入力ペアと類似する関係性を抽出することができる。Pantel らは入力としてある二項関係 (is-a など) を持つペア (cat, animal など) を与え、共起する文脈パターン (A **such as** B など) をもとに同様の関係性を有するペアを獲得する手法を提案している [21]。

このように、ブートストラップ法は様々なタスクに応用することから汎用的な手法である一方で、一回の繰り返し操作でいくつ出力結果を獲得するか、繰り返し操作をどれくらいするか、出力結果として獲得するための閾値をいくりに設定するか、など多くのパラメータ調整が必要となり、そのパラメータによって性能が大きく異なってくるため、ドメイン依存の手法でもある。また、繰り返し操作をしていく過程で無関係なシードを獲得した場合、手がかりとして曖昧なジェネリックパターンと呼ばれている共通パターンを獲得してしまい、結果として出力結果の精度が落ちるという意味ドリフトと呼ばれる問題が提起されており [19]、アルゴリズムによって意味ドリフトがいつどのように発生するのかという分析も行われている [20]。ブー

5.2. オンデマンドな知識獲得に関する研究

トストラップ法は *PMI* を用いたスコアリングによって共通パターンを獲得しているという点で，本研究における順序付け観点の獲得タスクと似ている点がある．しかしながら，本研究では比較文を利用したフィルタリングを行うことで出力インスタンスを制限するという，より順序付けタスクに特化した手法を提案しているため，これら手法とは異なっているといえる．

第6章 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するうえで重要であるということに着目し、複数の概念語を順序付けするタスクを提起した。提案手法では、類似する複数の概念語を多様な観点で順序付けしたいという問題に対して、複数の概念語に共通する順序付けの観点となる形容詞を獲得するタスクと、複数の概念語を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の2つに分け、それを実現した。

順序付けの観点となる形容詞の獲得タスクにおいては、比較表現によるフィルタリングを実装することで実装前と比べて適合率が向上していることが分かった。提案手法のほうが精度が高かったものの、実験結果でも述べたように、獲得された形容詞について分析をすると、形容詞の種類数はそこまで多くないことが分かる。この原因としては順序付けの観点として形容詞という一形態素だけは種類がそもそも多くないからだと考えられる。

今後の課題としては比較表現以外にも同位表現（～と～は同じくらい～だ）や、最上表現（～は～の中で最も～だ）などの文構造を用いて形容詞の絞込みを行うということで適合率を向上させることや二形態素以上からなる観点（経済が深刻だ、など）を獲得することが挙げられる。また、今回は適合率のみで評価を行ったが、被験者にカテゴリから想起される形容詞を挙げてもらうことで再現率についても検証する必要があると考えられる。

次に、概念語の順序付けタスクにおいては、対義形容詞を用いる提案手法が単一形容詞を用いる手法よりも相関係数が向上し、提案手法の有用性を示すことができた。誤り分析をした結果、概念語と形容詞との共起する文が概念語の性質を記述しているとは限らないということが分かった。より精度向上を図るにあたり現段階で

は2つの新たなアプローチを考えている。

まず1つ目は、共起頻度以外の手がかりを使うということである。現時点で考えられる案として、比較表現、同位表現、強調表現が挙げられる。比較表現とは、「～よりも」や「～に比べて」などのパターンである。同位表現とは「～と～は同じくらい～だ」などのパターンである。比較表現や同位表現や共起関係よりもより明確に概念語同士の性質の優劣に関して言及しているため、これらを考慮することにより、順序付けの誤りの原因になった、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。また、共起をカウントする際に「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強調表現の語を伴って共起をした場合の新たなスコア付けを考慮することで、よりよい結果になるのではにかと考えられる。

2つ目は、今回は順序付けの観点として形容詞の性質を利用し、概念語の順序付けを行ったが、例えばフレーズ（「給料の高さ」や「体長の長さ」）に拡張することで観点の曖昧性をなくし、順序付けの観点が明確化することで順序付けの相関を向上させることができるのではないかと考えている。

また、情報資源による結果の違いについても分析を行いたいと考えている。今回は共起頻度を獲得する際のテキストデータとしてブログ記事を用いたが、これをマイクロブログに変えて同実験を行い、結果を比較をすることで情報資源による結果の差異の議論も検討していきたい。

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 Journal*, Vol. 6, No. 1, 2007.
- [2] John R Taylor. *Linguistic Categorization*. Oxford: Clarendon Press, 1989.
- [3] P. F. Brown, V. J. Della Pietra, P. V. deSouza, J. C. Lai, and R. L. Mercer.
Class-based n -gram models of natural language *In Proceedings of Association
for Computational Linguistics*; vol. 18, pp. 467–479, 1992.
- [4] Yarowsky, D. Unsupervised Word Sense Disambiguation Rivaling Supervised
Methods. *In Proceedings of Association for Computational Linguistics*; pp.
189–196, 1995.
- [5] Nobuhiro Kaji and Masaru Kitsuregawa. Efficient Word Lattice Generation
for Joint Word Segmentation and POS tagging in Japanese. *In Proceedings of
IJCNLP*, October 2013.
- [6] Naoki Yoshinaga and Masaru Kitsuregawa. Kernel Slicing: Scalable Online
Training with Conjunctive Features. *In Proceedings of Journal*, pp. 1245–1253,
2010.
- [7] Naoki Yoshinaga and Masaru Kitsuregawa. Polynomial to Linear: Efficient
Classification with Conjunctive Features. *In Proceedings of Empirical Methods
in Natural Language Processing*; pp. 1542–1551, 2009.
- [8] Dekang Lin. An Information-Theoretic Definition of Similarity. *International
Conference on Machine Learning*, pp. 296–304, 1998.

-
- [9] Peter D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *In Proceedings of Association for Computational Linguistics*, pp. 417–424, July 2002.
- [10] Gerard de Melo and Mohit Bansal. Good, Great, Excellent: Global Inference of Semantic Intensities. *Transaction of Association for Computational Linguistics*, pp. 279–290, 2013.
- [11] Zoubin Ghahramani and Katherine A.Heller. Bayesian sets *In Advances in Neural Information Processing Systems*; 2005.
- [12] Richard C. Wang and William W. Cohen. Language-independent set expansion of named entities using the web. *In Proceedings of International Conference on Data Mining*; pp. 342–350, 2007.
- [13] Andrew Carlson, J. Betteridge, R.C. Wang, E.R. Hruschka Jr., and T.M. Mitchell. Language-independent set expansion of named entities using the web. *In Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM)*; 2012.
- [14] Yarowsky, D. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. *In Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*; pp. 189–1196, 1995.
- [15] Max Whitney and Anoop Sarkar Bootstrapping via Graph Propagation. *In Proceedings of Association for Computational Linguistics*; pp. 620–628,, 2012.
- [16] Collins, M. and Singer, Y. Unsupervised Models for Named Entity Classification *In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*; pp. 100–110, 1999.

-
- [17] Takaaki Hasegawa, Satoshi Sekine and Ralph Grishman. Discovering Relations among Named Entities from Large Corpora. *In Proceedings of Association for Computational Linguistics*; pp. 415–422, 2004
- [18] Nina TAHMASEBI, Gerhard GOSSEN, Nattiya KANHABUA, Helge HOLZMANN, Thomas RISSE. NEER: An Unsupervised Method for Named Entity Evolution Recognition. *In Proceedings of International Conference on Computational Linguistics*; pp. 2553–2568, 2012
- [19] Curran, J. R., Murphy, T., and Scholz, B. Minimising semantic drift with Mutual Exclusion Bootstrapping. *In Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*; pp. 172–180, 2007
- [20] 小町守, 工藤拓, 新保仁, 松本裕治. Espresso 型ブートストラッピング法における意味ドリフトのグラフ理論に基づく分析. *JSAI Journal* vol25. ver2, 2010
- [21] Patric Pantel and Marco Pennacchiotti. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. *In Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Associations for Computational Linguistics*; pp. 358–365, 2006.

発表文献

1. 仁科俊晴 , 鍛冶伸裕, 吉永直樹, 豊田正史. 対義形容詞対との相互情報量を利用した概念語の順序付け. 情報処理学会, 第 214 回自然言語処理研究会 (2013.10).
2. 仁科俊晴 , 吉永直樹, 鍛冶伸裕, 豊田正史. 形容詞との共起に基づく概念語の順序付け. Web とデータベースに関するフォーラム (WebDB Forum 2013) (2013.11) .
3. 仁科俊晴, 吉永直樹, 鍛冶伸裕, 豊田正史. 概念語に関する順序付け知識のオンデマンド獲得. 電子情報通信学会第 6 回データ工学と情報マネジメントに関するフォーラム／第 12 回日本データベース学会年次大会 (DEIM2014) (2014.03). (to appear)

付録 A 同一文内共起と係り受け関係の 順序付け結果の比較

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	ネコ	クマ	ネズミ	472	4023
2	イヌ	ネコ	ネコ	8084	82909
3	クマ	ネズミ	サル	448	4921
4	ウシ	サル	クマ	784	9212
5	キリン	ゾウ	ゾウ	225	3024
6	ゾウ	ライオン	ライオン	184	2505
7	サル	イヌ	イヌ	4619	67418
8	クジラ	キリン	キリン	157	3323
9	ライオン	ウシ	クジラ	126	3097
10	ネズミ	クジラ	ウシ	418	13425
相関係数	0.654	0.17	0.05		

表 6.1: 「可愛い」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	飛行機	新幹線	新幹線	314	12413
2	新幹線	飛行機	バス	294	31332
3	タクシー	バス	電車	329	46651
4	電車	自動車	飛行機	164	25801
5	バス	自転車	船	60	13982
6	船	船	自転車	187	44789
7	ヘリコプター	電車	タクシー	28	8624
8	自動車	タクシー	自動車	306	197989
9	自転車	ヘリコプター	ヘリコプター	1	1082
相関係数	0.777	0.355	0.385		

表 6.2: 「快適だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	飛行機	新幹線	新幹線	847	12413
2	新幹線	船	飛行機	275	25801
3	ヘリコプター	自転車	船	130	13982
4	電車	飛行機	自転車	366	44789
5	自動車	自動車	電車	315	46651
6	タクシー	ヘリコプター	ヘリコプター	7	1082
7	船	電車	自動車	705	197989
8	バス	バス	タクシー	29	8624
9	自転車	タクシー	バス	77	31332
相関係数	0.917	0.09	0.356		

表 6.3: 「速い」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	ワイン	マッコリ	マッコリ	331	1102
2	焼酎	ワイン	カクテル	1279	4509
3	カクテル	シャンパン	シャンパン	783	3061
4	ウイスキー	ビール	ビール	22597	95398
5	シャンパン	チューハイ	ワイン	8475	37127
6	ビール	カクテル	ハイボール	350	1559
7	ハイボール	ハイボール	チューハイ	152	951
8	チューハイ	焼酎	焼酎	1392	10417
9	マッコリ	ウイスキー	ウイスキー	217	2508
相関係数	0.528	-0.173	-0.144		

表 6.4: 「美味しい」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	寿司	ピザ	ピザ	6757	15141
2	ステーキ	パスタ	パスタ	10860	27578
3	ピザ	チャーハン	チャーハン	2499	6945
4	ラーメン	ステーキ	ステーキ	2019	5909
5	カレー	パン	パン	28047	86954
6	パスタ	寿司	ハンバーガー	1769	6152
7	ハンバーガー	カレー	寿司	5113	17897
8	チャーハン	焼きそば	焼きそば	1984	7048
9	焼きそば	ハンバーガー	カレー	15699	59895
10	パン	ラーメン	ラーメン	11773	56404
相関係数	0.733	-0.001	-0.108		

表 6.5: 「美味しい」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	ハンバーガー	ステーキ	ステーキ	6	5909
2	カレー	寿司	寿司	16	17897
3	パン	パスタ	ピザ	11	15141
4	焼きそば	ピザ	パスタ	16	27578
5	チャーハン	チャーハン	ハンバーガー	2	6152
6	ラーメン	焼きそば	ラーメン	18	56404
7	パスタ	ラーメン	焼きそば	2	7048
8	ピザ	ハンバーガー	パン	22	86954
9	寿司	カレー	カレー	14	59895
10	ステーキ	パン	チャーハン	1	6945
相関係数	0.796	-0.757	-0.651		

表 6.6: 「リーズナブルだ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	野球	野球	ラグビー	5	2594
2	サッカー	ラグビー	ボクシング	6	3900
3	水泳	プロレス	水泳	5	4594
4	相撲	ボクシング	サッカー	46	48087
5	バレー	サッカー	バレー	4	4455
6	テニス	水泳	プロレス	6	8129
7	ゴルフ	バレー	テニス	5	11476
8	ラグビー	ゴルフ	野球	16	45310
9	ボクシング	テニス	相撲	2	7928
10	プロレス	相撲	ゴルフ	2	20127
相関係数	0.888	-0.23	-0.371		

表 6.7: 「メジャーだ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	サッカー	プロレス	野球	4903	45310
2	野球	野球	相撲	847	7928
3	テニス	相撲	ラグビー	256	2594
4	水泳	サッカー	ボクシング	315	3900
5	バレー	ラグビー	サッカー	3743	48087
6	ゴルフ	ボクシング	プロレス	601	8129
7	ラグビー	バレー	バレー	271	4455
8	ボクシング	テニス	テニス	666	11476
9	相撲	ゴルフ	ゴルフ	1081	20127
10	プロレス	水泳	水泳	71	4594
相関係数	0.773	-0.232	-0.077		

表 6.8: 「面白い」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	レモン	レモン	グレープフルーツ	39	1594
2	グレープフルーツ	グレープフルーツ	レモン	72	4742
3	パイナップル	ミカン	イチゴ	311	30060
4	ミカン	パイナップル	パイナップル	27	2612
5	イチゴ	イチゴ	ミカン	84	10700
6	マンゴー	リンゴ	マンゴー	32	7471
7	ナシ	マンゴー	リンゴ	65	16625
8	リンゴ	ナシ	ナシ	5	5675
9	カキ	モモ	モモ	13	18341
10	モモ	カキ	カキ	2	10684
相関係数	0.828	0.786	0.763		

表 6.9: 「酸っぱい」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	ナシ	マンゴー	マンゴー	1728	7471
2	モモ	パイナップル	ナシ	1230	5675
3	パイナップル	イチゴ	パイナップル	542	2612
4	イチゴ	カキ	リンゴ	2923	16625
5	リンゴ	リンゴ	カキ	1867	10684
6	マンゴー	グレープフルーツ	ミカン	1868	10700
7	ミカン	レモン	イチゴ	4796	30060
8	グレープフルーツ	ナシ	グレープフルーツ	231	1594
9	カキ	ミカン	モモ	1914	18341
10	レモン	モモ	レモン	339	4742
相関係数	0.741	-0.113	0.278		

表 6.10: 「美味しい」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	アメリカ	フランス	イギリス	77	5555
2	中国	イギリス	フランス	75	7964
3	イギリス	イタリア	ドイツ	84	9260
4	フランス	ドイツ	カナダ	23	2573
5	イタリア	アメリカ	イタリア	51	7703
6	ドイツ	カナダ	インド	28	4363
7	ロシア	ブラジル	ブラジル	30	4754
8	インド	インド	ロシア	17	4676
9	カナダ	ロシア	中国	55	23212
10	ブラジル	中国	アメリカ	61	29166
相関係数	0.818	0.22	-0.141		

表 6.11: 「有名だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	インド	インド	ロシア	5	4676
2	中国	アメリカ	インド	3	4363
3	ブラジル	イギリス	中国	11	23212
4	カナダ	ロシア	アメリカ	13	29166
5	イタリア	中国	ドイツ	4	9260
6	フランス	フランス	イタリア	3	7703
7	ロシア	イタリア	カナダ	1	2573
8	イギリス	ブラジル	フランス	2	7964
9	アメリカ	ドイツ	ブラジル	1	4754
10	ドイツ	カナダ	イギリス	1	5555
相関係数	0.585	0.073	0.055		

表 6.12: 「貧乏だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	札幌	札幌	札幌	2471	15682
2	仙台	仙台	仙台	1537	10526
3	京都	沖縄	福岡	1105	10490
4	大阪	京都	京都	2089	26495
5	名古屋	福岡	大阪	2085	31189
6	横浜	大阪	沖縄	1255	19739
7	神戸	横浜	名古屋	922	17454
8	福岡	名古屋	横浜	672	13119
9	沖縄	神戸	神戸	389	8371
相関係数	0.917	0.251	0.375		

表 6.13: 「寒い」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	ネックレス	指輪	ブレスレット	2	1749
2	指輪	ネックレス	指輪	6	5567
3	ピアス	ブレスレット	イヤリング	1	959
4	ブレスレット	イヤリング	ネックレス	5	5121
5	イヤリング	ピアス	ピアス	2	5635
相関係数	0.478	0.336	-0.311		

表 6.14: 「高価だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	パソコン	カーナビ	ビデオカメラ	7	1740
2	ビデオカメラ	ビデオカメラ	カーナビ	10	3981
3	カメラ	プリンター	プリンター	12	9523
4	カーナビ	カメラ	パソコン	100	121580
5	スマートフォン	スマートフォン	カメラ	60	73180
6	プリンター	パソコン	スマートフォン	5	6296
相関係数	0.825	-0.276	0.042		

表 6.15: 「高価だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	ファミコン	ファミコン	ファミコン	2	1220
2	W i i	D S	W i i	4	5450
3	D S	P S P	D S	5	8997
4	P S P	W i i	P S P	3	6214
相関係数	0.355	0.25	0.355		

表 6.16: 「しょぼい」に基づく順序づけ

順位	オラクル	文内共起	概念語	係り受け 共起頻度	概念語頻度
1	W i i	ファミコン	W i i	223	5450
2	D S	W i i	ファミコン	45	1220
3	ファミコン	D S	D S	196	8997
4	P S P	P S P	P S P	57	6214
相関係数	0.25	0.05	0.2		

表 6.17: 「面白い」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	W i i	W i i	P S P	55	6214
2	D S	P S P	D S	35	8997
3	P S P	D S	ファミコン	4	1220
4	ファミコン	ファミコン	W i i	6	5450
相関係数	0.948	0.743	-0.384		

表 6.18: 「新しい」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	グーグル	グーグル	アマゾン	569	5934
2	アマゾン	ヤフー	グーグル	152	4239
3	ヤフー	アマゾン	ヤフー	25	2011
相関係数	1	0.5	0.5		

表 6.19: 「便利だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	グーグル	ヤフー	ヤフー	3	2011
2	ヤフー	グーグル	グーグル	5	4239
3	アマゾン	アマゾン	アマゾン	4	5934
相関係数	0.5	0.25	0.25		

表 6.20: 「簡単だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	グーグル	グーグル	グーグル	21	4239
2	アマゾン	アマゾン	ヤフー	4	2011
3	ヤフー	ヤフー	アマゾン	1	5934
相関係数	0.75	0.75	0.75		

表 6.21: 「巨大だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	サクラ	モミジ	モミジ	31685	77074
2	モミジ	アジサイ	サクラ	33482	109533
3	アジサイ	イチョウ	アジサイ	4996	16402
4	ウメ	コスモス	コスモス	1826	6277
5	イチョウ	サクラ	イチョウ	1830	7165
6	ヒマワリ	チューリップ	チューリップ	1243	6580
7	コスモス	ウメ	ウメ	2073	16528
8	チューリップ	タンポポ	タンポポ	195	2414
9	タンポポ	ヒマワリ	ヒマワリ	386	6606
相関係数	0.806	0.35	0.517		

表 6.22: 「綺麗だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	競馬	競艇	競艇	208	846
2	競艇	パチスロ	パチスロ	235	992
3	パチンコ	パチンコ	パチンコ	1224	5231
4	パチスロ	競馬	競馬	3356	18247
相関係数	0.55	-0.35	-0.35		

表 6.23: 「好きだ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	競馬	競艇	競艇	87	846
2	競艇	パチスロ	競馬	1247	18247
3	パチンコ	パチンコ	パチスロ	48	992
4	パチスロ	競馬	パチンコ	202	5231
相関係数	0.55	-0.35	0.25		

表 6.24: 「面白い」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	競馬	パチスロ	競艇	11	846
2	競艇	競艇	パチスロ	8	992
3	パチンコ	競馬	競馬	99	18247
4	パチスロ	パチンコ	パチンコ	9	5231
相関係数	0.495	-0.332	-0.068		

表 6.25: 「簡単だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	アイスクリーム	アップルパイ	アップルパイ	1020	2695
2	チョコレート	アイスクリーム	アイスクリーム	9403	29470
3	ポテトチップス	ホットケーキ	ホットケーキ	638	2220
4	ホットケーキ	チョコレート	ポップコーン	374	1417
5	アップルパイ	ポテトチップス	ポテトチップス	486	2914
6	ポップコーン	ポップコーン	チョコレート	8761	52851
相関係数	0.793	0.206	-0.139		

表 6.26: 「美味しい」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	アイスクリーム	アップルパイ	ポテトチップス	310	2914
2	チョコレート	ポテトチップス	チョコレート	5193	52851
3	ポテトチップス	チョコレート	ホットケーキ	202	2220
4	ホットケーキ	ホットケーキ	アップルパイ	204	2695
5	アップルパイ	アイスクリーム	アイスクリーム	2032	29470
6	ポップコーン	ポップコーン	ポップコーン	80	1417
相関係数	0.807	-0.139	0.105		

表 6.27: 「大好きだ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	バースデー	クリスマス	クリスマス	239	29282
2	クリスマス	バレンタインデー	バレンタインデー	45	9617
3	バレンタインデー	ホワイトデー	ハロウィン	23	5758
4	ハロウィン	バースデー	バースデー	161	47268
5	ホワイトデー	ハロウィン	イースター	2	601
6	イースター	イースター	ホワイトデー	3	1814
相関係数	0.702	0.294	0.307		

表 6.28: 「素敵だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	政治	数学	数学	2866	17091
2	音楽	地理	地理	493	3042
3	歴史	古典	古典	145	2429
4	古典	歴史	歴史	1353	52016
5	地理	音楽	音楽	495	142527
6	数学	政治	政治	41	16410
相関係数	0.835	-0.9	-0.9		

表 6.29: 「苦手だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	数学	歴史	政治	79	16410
2	政治	地理	古典	7	2429
3	地理	政治	地理	8	3042
4	歴史	数学	音楽	346	142527
5	音楽	古典	数学	36	17091
6	古典	音楽	歴史	86	52016
相関係数	0.712	0.213	-0.236		

表 6.30: 「重要だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	ソフトバンク	ライオンズ	カープ	72	3810
2	ライオンズ	カープ	ライオンズ	23	1287
3	カープ	ソフトバンク	ソフトバンク	117	8364
4	ベ이스ターズ	ベ이스ターズ	ベ이스ターズ	15	1196
相関係数	0.898	0.387	0.176		

表 6.31: 「好調だ」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	ソフトバンク	カープ	ライオンズ	211	1287
2	ライオンズ	ライオンズ	カープ	575	3810
3	カープ	ベ이스ターズ	ソフトバンク	930	8364
4	ベ이스ターズ	ソフトバンク	ベ이스ターズ	132	1196
相関係数	0.924	-0.395	0.369		

表 6.32: 「強い」に基づく順序づけ

順位	オラクル	文内共起	係り受け		
			概念語	共起頻度	概念語頻度
1	ベ이스ターズ	カープ	ベ이스ターズ	7	1196
2	ライオンズ	ベ이스ターズ	ライオンズ	7	1287
3	カープ	ライオンズ	ソフトバンク	20	8364
4	ソフトバンク	ソフトバンク	カープ	9	3810
相関係数	0.724	0.313	0.571		

表 6.33: 「酷い」に基づく順序づけ

付録B 単一形容詞と対義形容詞対の 順序付け結果の比較

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	飛行機	新幹線	船	1671	1671	499619
2	新幹線	船	ヘリコプター	74	85	38516
3	ヘリコプター	自転車	新幹線	2975	3710	388588
4	電車	飛行機	自転車	3691	5106	1178490
5	自動車	自動車	自動車	11133	23242	4659037
6	タクシー	ヘリコプター	飛行機	1389	3333	571667
7	船	電車	電車	2907	16326	1592251
8	バス	バス	バス	1125	7884	952336
9	自転車	タクシー	タクシー	357	3039	343171
相関係数	0.917	0.09	0.082			

表 6.34: 「速い」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	野球	野球	野球	1435	309	565904
2	サッカー	ラグビー	ゴルフ	162	38	349583
3	水泳	プロレス	バレー	39	19	71505
4	相撲	ボクシング	ボクシング	61	30	78635
5	バレー	サッカー	テニス	91	48	207772
6	テニス	水泳	プロレス	107	62	106456
7	ゴルフ	バレー	サッカー	536	341	726470
8	ラグビー	ゴルフ	水泳	42	27	73246
9	ボクシング	テニス	相撲	28	20	126886
10	プロレス	相撲	ラグビー	74	84	45584
相関係数	0.888	-0.23	0.044			

表 6.35: 「メジャーだ」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	サッカー	プロレス	水泳	469	24	73246
2	野球	野球	テニス	2999	159	207772
3	テニス	相撲	プロレス	3190	196	106456
4	水泳	サッカー	バレー	1058	66	71505
5	バレー	ラグビー	ゴルフ	3970	331	349583
6	ゴルフ	ボクシング	ラグビー	1049	102	45584
7	ラグビー	バレー	野球	14678	1503	565904
8	ボクシング	テニス	相撲	3109	343	126886
9	相撲	ゴルフ	ボクシング	1619	197	78635
10	プロレス	水泳	サッカー	17279	2525	726470
相関係数	0.773	-0.232	-0.058			

表 6.36: 「面白い」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	インド	インド	アメリカ	349	621	1498683
2	中国	アメリカ	中国	257	488	1349491
3	ブラジル	イギリス	インド	58	142	238976
4	カナダ	ロシア	ロシア	45	113	230867
5	イタリア	中国	ブラジル	17	47	140480
6	フランス	フランス	カナダ	10	30	126415
7	ロシア	イタリア	フランス	56	169	411376
8	イギリス	ブラジル	イタリア	43	141	320958
9	アメリカ	ドイツ	イギリス	67	229	305431
10	ドイツ	カナダ	ドイツ	37	167	379021
相関係数	0.585	0.073	0.204			

表 6.37: 「貧乏だ」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	札幌	札幌	仙台	5758	3326	296983
2	仙台	仙台	札幌	11664	7272	389882
3	京都	沖縄	福岡	4811	3250	387837
4	大阪	京都	横浜	4792	3541	486001
5	名古屋	福岡	神戸	2442	1980	325283
6	横浜	大阪	京都	10750	12564	859176
7	神戸	横浜	沖縄	8538	10318	673929
8	福岡	名古屋	大阪	11620	15401	1172908
9	沖縄	神戸	名古屋	5072	8330	539679
相関係数	0.917	0.251	0.311			

表 6.38: 「寒い」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	ネックレス	指輪	指輪	541	19	118760
2	指輪	ネックレス	ネックレス	352	32	136302
3	ピアス	ブレスレット	ピアス	156	21	115624
4	ブレスレット	イヤリング	ブレスレット	93	18	53511
5	イヤリング	ピアス	イヤリング	38	10	22898
相関係数	0.478	0.336	0.429			

表 6.39: 「高価だ」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	パソコン	カーナビ	カメラ	2269	856	1734820
2	ビデオカメラ	ビデオカメラ	ビデオカメラ	103	47	36676
3	カメラ	プリンター	カーナビ	165	98	56054
4	カーナビ	カメラ	プリンター	194	174	135146
5	スマートフォン	スマートフォン	パソコン	1741	1896	2896229
6	プリンター	パソコン	スマートフォン	127	205	182971
相関係数	0.825	-0.276	0.087			

表 6.40: 「高価だ」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	W i i	ファミコン	W i i	3141	83	128828
2	D S	W i i	P S P	2668	73	160562
3	ファミコン	D S	D S	4904	185	219423
4	P S P	P S P	ファミコン	867	33	34661
相関係数	0.25	0.05	0.2			

表 6.41: 「面白い」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	W i i	W i i	W i i	2945	289	128828
2	D S	P S P	D S	3631	648	219423
3	P S P	D S	P S P	2930	587	160562
4	ファミコン	ファミコン	ファミコン	347	412	34661
相関係数	0.948	0.743	0.948			

表 6.42: 「新しい」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	グーグル	グーグル	グーグル	3408	119	209431
2	アマゾン	ヤフー	アマゾン	3065	151	390473
3	ヤフー	アマゾン	ヤフー	1301	102	156998
相関係数	1	0.5	1			

表 6.43: 「便利だ」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	グーグル	ヤフー	ヤフー	2149	65	156998
2	ヤフー	グーグル	アマゾン	2524	167	390473
3	アマゾン	アマゾン	グーグル	2695	187	209431
相関係数	0.5	0.25	-0.25			

表 6.44: 「簡単だ」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	競馬	競艇	競艇	550	16	18519
2	競艇	パチスロ	競馬	8560	321	425481
3	パチンコ	パチンコ	パチスロ	664	38	23904
4	パチスロ	競馬	パチンコ	3467	404	126980
相関係数	0.55	-0.35	0.25			

表 6.45: 「好きだ」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	競馬	競艇	競艇	297	13	18519
2	競艇	パチスロ	競馬	5486	494	425481
3	パチンコ	パチンコ	パチンコ	1698	190	126980
4	パチスロ	競馬	パチスロ	371	44	23904
相関係数	0.55	-0.35	0.35			

表 6.46: 「面白い」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	競馬	パチスロ	競艇	220	9	18519
2	競艇	競艇	競馬	3228	232	425481
3	パチンコ	競馬	パチンコ	717	97	126980
4	パチスロ	パチンコ	パチスロ	294	43	23904
相関係数	0.495	-0.332	0.395			

表 6.47: 「簡単だ」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	政治	数学	歴史	3480	991	1305642
2	音楽	地理	古典	526	207	97212
3	歴史	古典	地理	1151	516	71355
4	古典	歴史	音楽	3493	1880	2638372
5	地理	音楽	数学	8563	5099	239859
6	数学	政治	政治	621	646	1195411
相関係数	0.835	-0.9	-0.153			

表 6.48: 「苦手だ」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	ソフトバンク	ライオンズ	ソフトバンク	589	257	191272
2	ライオンズ	カープ	カープ	197	109	62464
3	カープ	ソフトバンク	ライオンズ	104	74	28647
4	ベ이스ターズ	ベ이스ターズ	ベ이스ターズ	58	68	24678
相関係数	0.898	0.387	0.687			

表 6.49: 「好調だ」に基づく順序づけ

順位	オラクル	ベースライン	提案手法			
			概念語	共起	反義共起	出現頻度
1	ソフトバンク	カープ	ライオンズ	863	245	28647
2	ライオンズ	ライオンズ	ソフトバンク	4072	1584	191272
3	カープ	ベ이스ターズ	カープ	2251	901	62464
4	ベ이스ターズ	ソフトバンク	ベ이스ターズ	679	409	24678
相関係数	0.924	-0.395	0.724			

表 6.50: 「強い」に基づく順序づけ