

多様な手がかりを用いた形容詞に基づく概念語の順序付け

岩成 達哉[†] 吉永 直樹^{††,†††} 豊田 正史^{††} 喜連川 優^{††,††††}

[†] 東京大学大学院 情報理工学系研究科 〒113-8654 東京都文京区本郷 7-3-1

^{††} 東京大学 生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

^{†††} 情報通信研究機構 〒184-8795 東京都小金井市貫井北町 4-2-1

^{††††} 国立情報学研究所 〒101-0003 東京都千代田区一ツ橋 2-1-2

E-mail: †{nari,ynaga,toyoda,kitsure}@tkl.iis.u-tokyo.ac.jp

あらまし 本稿では、複数の概念語 (例: ロンドン, ハワイ, ローマ) を共通する性質 (例: 安全だ) の強さで順序付けするタスクに取り組む. 具体的には, 大規模ウェブテキストから概念語と形容詞の共起や比喻表現などの様々な順序付けの手がかりを抽出し, これらを特徴量として教師ありランキング学習により順序付けを行う. 実験では, 我々の研究室が有するブログ記事や Twitter への投稿を解析して手がかりを集め, 客観的・主観的評価に基づく複数の概念語と形容詞の組について, 提案手法と人手による順序付けとの相関を調べることで, 提案手法の有用性を検証した. キーワード 自然言語処理, ソーシャルメディア, ランキング学習, 概念語の順序付け

1. はじめに

日常生活において, 我々は様々な観点で物事に順序をつけ, 利用するものを選んでいく. 例えば, 遠い目的地に早くたどり着きたいのなら, 徒歩よりも自転車, 自転車よりも車, 車よりも飛行機を利用する. このとき, 我々は「車」には個体差があることは知りながらも, 典型的な個体を思い浮かべて「車」という概念を捉え, このような順序付けを行っている. 上記のような事物の順序に関する知識は意思決定を行う上で非常に有用であるが, 「お寺」を「古い」順に並べ替えるなどのように容易に判断ができない問題も少なくなく, そのための情報収集に大きく時間をとられてしまう.

仁科ら [20] は, このような概念語 (徒歩, 車, 飛行機) を, 共通する性質の強さ (速い) に基づいて順序付けするタスクを提案した. このタスクでは, クエリとして, 調べたい性質を表す形容詞と共に, 概念語の集合が与えられ, 出力として順序付けされた概念語のリストが得られる. 例えば, クエリとして形容詞「速い」が概念語集合 {自転車, 車, 飛行機} と共に与えられた場合は, 出力として飛行機 > 車 > 自転車が期待される.

タスクを定義する際に問題となるのは, 「アニメ」を「面白い」順に並び替えるような主観的な順序付けにおいて, 正解をどのように設定するかということである. 本稿では, 速度などの数値に紐付けられ客観的に順序付け可能な問題については, 実際の数値を基にした順序付けを正解とする一方で, 主観的な順序付けについては, 仁科らにならい, 概念語の考えられるすべての順序に対して, 複数の人手による順序付けとのスピアマンの順位相関係数 ρ を計算し, その平均が最大となる概念語の順序を正解とする. カフェの好みのように主観的評価に依存する問題では絶対的な正解は必ずしも存在しないが, これによって得られた解は, 複数人の順序付けに関する認識を最大公約数的に反映したものとなり, テキストから概念語の順序に関する共通認識を得ようとする本研究の目的に見合ったものとなっている.

仁科らは, このタスクに対して, 大量のウェブテキストを解析し, 形容詞と概念語の文内共起や, 概念語から形容詞への係り受けの出現頻度を数え上げ, それぞれの頻度を基に順序付けを行う手法を提案した. 本研究では, 既存手法で用いられていた手がかりである, 形容詞と概念語の共起や係り受けだけでなく, 直喩表現や比較表現といった手がかりを取り入れ, 教師ありランキング学習の枠組みで, 特徴量として同時に考慮して順序付けを行う. 本稿では, 教師ありランキング学習の手法として順序付け対象の任意の 2 つの要素の組について, 正解の順序に対する順序関係の誤りを最小化するランキング SVM (Ranking Support Vector Machine) [7] と相関係数を目的変数とした SVR (Support Vector Regression) [6] を利用し, それぞれの比較を行う.

実験では, 我々の研究室が有する 8 年分の日本語ブログ記事や 3 年分の Twitter への投稿を解析して手がかりを集め, ブログ記事を基にして作成した客観的・主観的評価に基づく形容詞と概念語集合の組に対して, 提案手法が出力する順序と, 人手による順序付けを基に得られた正解の順序との相関係数を求めることで手法の有用性を確かめる.

2. 関連研究

本研究で扱う, 形容詞が表す性質の強さで概念語を順序付けるタスクは, 仁科ら [20] によって提案されたものである. 仁科らは, 大量のウェブテキストを解析し, 概念語と形容詞の文内共起や, 概念語から形容詞への係り受けを数え上げ, それらの頻度を基に, 自己相互情報量 PMI (Point-wise Mutual Information) を計算して, 順序付けを行っている. 仁科らはこれらの手がかりをそれぞれ独立に利用して概念語の順序付けを行っており, 手がかりを組み合わせて使っていない. 一方, 本研究では, 教師ありランキング学習を用い, これらの手がかりを特徴量として学習を行うことで, 多様な手がかりを組み合わせることを可能とした.

質問応答システムは様々なテキストを解析することによって、現実の問題 (例: 「東京の平均気温は?」) に対する回答を出力する [12]。このような、ある事物の特徴やその値をウェブテキストから求める研究は複数ある [1], [4], [14], [17], [19]。Yoshinaga と Torisawa [19] はウェブの表や箇条書きから事物の属性と値を抽出する手法を提案している。また, Takamura と Tsujii [14] は, 事物の数値に基づく特徴 (例: 大きさ, 重さ) を推定するタスクに対し, 事物の特徴に関する比較情報などの手がかりを集める手法を提案している。これらの研究は, 特に数値によって表すことのできる性質に基づく順序付けに利用可能である。

事物に対する感想や感情の分析では, ウェブ上のレビューや意見などを解析し, 製品に関する評価を集める [11]。Kurashima ら [9] は, 比較表現を集めて, 事物の人気などを順序付けする手法を提案している。この手がかりは本研究でも利用しているが, 本研究では比較表現を様々な情報と組み合わせ, 直接的に言及されていない事物間の順序も考慮している点で異なる。

また, 部分的な順序の関係をまとめて順序付けを行う研究もある [2], [5], [10], [13], [16]。これらの研究では, 2 つの事物に関する順序が複数与えられた時に, それらの情報を組み合わせて元の問題を解く。一方で, 本タスクでは, 概念語間の直接的な比較表現が頻出しなことから, 手法の汎用性を保つために部分順序が与えられることを前提としておらず, 他の特徴量と組み合わせで扱っている。

これらの研究に対し, 順序付けという観点でみると数値などで表せる客観的な順序付けだけではなく, 主観的な順序付けも行い, また具体物だけでなく概念語も扱うため, 本稿で取り上げているタスクの方がより一般的なタスクであるといえる。

3. 提案手法

本研究では, ソーシャルメディアテキストを解析することで, 3.1 節に示す多様な手がかりを集める。これらの手がかりは, 我々が持っている概念語に対する認識は, 我々が書く文章に明示的あるいは暗黙的に影響するという仮定に則って定めている。このようにして集めた手がかりを特徴量として教師ありランキング学習に用いることで, 概念語を形容詞に基づいて並び替える (3.2 節)。

3.1 順序付けに用いる手がかり

本手法では, 仁科らが提案した形容詞と概念語の文内共起, 概念語から形容詞への係り受けだけでなく, 直喩表現と比較表現を含めた 4 つの手がかりを順序付けに利用する。本研究では, 日本語のクエリを用いて評価を行うが, ここで示す手がかりは言語に非依存なものであり, 容易に他言語への拡張が可能である。最初の 3 つは, 暗黙的に概念語の性質の強さを示すが, 4 つ目の手がかりは, 明示的に 2 つ以上の概念語間の部分的な順序関係を示すものである。

概念語と形容詞の文内共起 概念語がある性質を持つのであれば, 我々はその性質を表す形容詞を概念語と一緒に述べる事が多い (例: クジラ はなんて 大きい んだ)。したがって, 概念語と形容詞の文内共起を数えることで, 概念語の性質の強さを推定できる。

概念語から形容詞への係り受け 概念語から形容詞への係り受けは, 概念語の性質を直接的に示している。この手がかりは文内共起よりも頻度が少なくなるが, 文内共起と異なり係り受けでは概念語と形容詞の関係を直接的に示すため (例: アリ は ゾウ に踏まれないくらい 小さい) , より有用な手がかりになると考えられる。

直喩表現 我々はその性質の強さが非常に大きい概念語を喩えとして明示的に取り上げることがある (例: まるで 雪 のように 白い)。実験では, このような直喩表現のパターンを列挙し, それらにマッチするものを数え上げた。

比較表現 比較表現は, 複数の概念語間の関係を直接的に表す表現である (例: ロシア は 中国 よりも 寒い)。実験では, 直喩表現と同様に, 比較表現のパターンを列挙し, それらにマッチするものを数え上げた。

本研究では, 与えられた形容詞に対して対義語を 1 つ選び, 与えられた形容詞と同様に上記の手がかりを集めることで, 性質に関する負の極性の情報も集める。さらに, それらの形容詞に否定が付属する場合は, それぞれ反対の極性の手がかりとして数え上げている (例: 「イヌは大きくない」は「イヌは小さい」と同様と考える)^(注1)。

3.2 多様な手がかりを用いた教師ありランキング学習による概念語の順序付け

本研究では, 3.1 節で示した多様な手がかりを順序付けに利用するために, 教師ありランキング学習手法として, ランキング SVM (Ranking Support Vector Machine) [7] と SVR (Support Vector Regression) [6] を利用し, 比較を行う。

3.2.1 ランキング SVM

ランキング SVM は, 手がかり (特徴量) を基に, 順序付け対象のそれぞれ 2 つの要素の組について, 正解の順序と比較した部分順序の誤りの数を最小化するように特徴量に対する重みの学習を行う。

クエリを変えると概念語の数や単語の出現頻度が異なるため, 本稿では学習が正しく行えるようにそれぞれの特徴量を正規化している。3.1 節で取り上げた手がかりのうち, 最初の 3 つの手がかりは, 与えられた形容詞 (またはその対義語の否定) だけでなく, 与えられた形容詞の対義語 (または与えられた形容詞の否定) についても考慮し, それぞれの自己相互情報量 PMI を式 (1) で計算する。

$$PMI(x, y) = \log \frac{p(x, y)}{p(x) \cdot p(y)} \quad (1)$$

このとき, 形容詞とその対義形容詞について, 正と負の 2 つの極性をもつ PMI が得られることになるが, 仁科らにならって Turney ら [15] の評価極性の分類に基づき, 2 つの極性を 1 つにまとめて特徴量とする。例えばある形容詞との文内共起に関するある概念語の特徴量 $\phi(x)$ 文内共起 (ただし, $x = (\text{概念語}, \text{形容詞})$)

(注1): 直喩表現においては, 付属する否定がその性質の否定を表すわけではないため, 否定が付属するかどうかは調べていない (例: 「それは雪みたいに白くない」は雪が白いことを表す)。

表 1 ランキング SVM で用いた正規化された特徴量 (概念語ごと)

特徴量の種類	特徴量
文内共起	$\phi(x)$ 文内共起
係り受け	$\phi(x)$ 係り受け
直喩表現	$\phi(x)$ 直喩表現
比較表現 (正)	優っていると記述された他の概念語数/他の概念語数
比較表現 (負)	劣っていると記述された他の概念語数/他の概念語数

は、式 (2) のようになる^(注2)。

$$\begin{aligned} \phi(x)_{\text{文内共起}} & \quad (2) \\ &= SO_{\text{形容詞}}^{\text{形容詞}}(\text{概念語}) \\ &= \text{PMI}(\text{概念語}, \text{肯定の形容詞} [\text{形容詞} \text{ or } \text{対義語の否定}]) \\ &\quad - \text{PMI}(\text{概念語}, \text{否定の形容詞} [\text{対義語} \text{ or } \text{形容詞の否定}]) \\ &= \log \frac{p(\text{概念語}, \text{肯定の形容詞}) \cdot p(\text{否定の形容詞})}{p(\text{概念語}, \text{否定の形容詞}) \cdot p(\text{肯定の形容詞})} \end{aligned}$$

4 つ目の手がかりは、2 つの概念語の間の比較表現である。この手がかりは、ある概念語が、与えられた他の概念語のうち何個に対して、コーパス内で 1 度でも優っていると記述されたかを特徴量とする。例えば、クエリとして概念語集合 {イヌ, ゾウ, クジラ} と形容詞「大きい」が与えられ、「クジラはゾウよりも大きい」と「クジラはイヌよりも大きい」のいずれもがコーパス内に 1 回以上出現した場合は、クジラに関してこの特徴量は 2 (ゾウとイヌの 2 つの概念語に優っている記述があった) となる。また、反対の極性の特徴量として、与えられた他の概念語のうち、何個に対して 1 度でも劣っていると記述されたか (上記の場合は、ゾウ・イヌ共にクジラに劣っていると記述されたか) と考える) を同様にして利用する。これらの値は、与えられた概念語の数によって正規化する。

以上をまとめると、ランキング SVM では、各概念語について表 1 のような特徴量をコーパスから抽出し、各概念語の評価値を求め、その大きさの順に概念語を並べることで順序付けを行う。

3.2.2 SVR

3.2.1 節で示したランキング SVM では概念語ごとに特徴量を選ぶが、SVR では概念語の順序ごとに特徴量を選ぶ手法をとる。SVR は回帰分析を行う手法であるが、本稿では概念語の考えられるすべての順序と正解の順序とのスピアマンの相関係数を計算し、それを目的変数として、それぞれの特徴量への重みを学習する。テストでは提案手法による順序と、正解の順序とのスピアマンの相関係数を用いるために、この回帰によって、どの順序が最も良いと考えられるかを直接的に選択できる。

ところで、SVR では、考えられるすべての順序を学習の教師データとして利用できるが、それらの順序の数は概念語の数を n とすると $n!$ となり、クエリの概念語の数によって学習に用いる教師データの数が偏ってしまう。そこで本稿では、ある概念語集合と形容詞の組に特化した学習が行われないように、すべてのクエリの概念語集合の中で最も少ない数を n_{min} とし

表 2 SVR で用いた正規化された特徴量 (順序ごと)

特徴量の種類	特徴量 (順序の任意の 2 個の概念語について)
文内共起	$f_{\text{文内共起}}^{\text{SVM}}$ の大小関係が正解と同順の数 $/N C_2$
係り受け	$f_{\text{係り受け}}^{\text{SVM}}$ の大小関係が正解と同順の数 $/N C_2$
直喩表現	$f_{\text{直喩表現}}^{\text{SVM}}$ の大小関係が正解と同順の数 $/N C_2$
比較表現 (正)	$f_{\text{比較表現 (正)}}^{\text{SVM}}$ の大小関係が正解と同順の数 $/N C_2$
比較表現 (負)	$f_{\text{比較表現 (負)}}^{\text{SVM}}$ の大小関係が正解と逆順の数 $/N C_2$

たとき、それぞれのクエリからは同じ数 $n_{min}!$ だけ順序を取り出して教師データとしている。一方、テストでは、すべての順序についてのスピアマンの相関係数を回帰によって推定し、最も良いものを選ぶ。

学習に利用する順序をサンプリングした後、取り出したそれぞれの順序に対して正解の順序とのスピアマンの相関係数 (目的変数) と、特徴量 (説明変数) を求めて、特徴量への重みを学習する。特徴量には、作成した順序から任意の 2 個を選び、表 1 に示した各特徴量の大きさの大小関係が、その順序において正しく成り立っているかを数え上げ、任意の 2 個の概念語を取り出す場合の数で割って正規化して用いている。例えば、形容詞「大きい」において、クジラとイヌの係り受けに関する特徴量の大小関係が $SO_{\text{係り受け}}^{\text{大きい}}(\text{クジラ}) > SO_{\text{係り受け}}^{\text{大きい}}(\text{イヌ})$ であるならば、取り出した順序でクジラ $>$ イヌのときに正しく成り立っており、逆にイヌ $>$ クジラのときに成り立っていないとする。

表 1 の最初の 4 つの特徴量については、以上のようにして特徴量を求め、最後の比較表現 (負) に関しては、負の極性を表す値であるために大小関係を反対にした判定を行って特徴量とする。各クエリにおいて、これら 5 つの特徴量を順序ごとに求め、相関係数とともに学習・テストに利用する。

以上をまとめると、SVR で用いた特徴量はランキング SVM で用いた特徴量を $f_{\text{文内共起}}^{\text{SVM}}$ のように表すとすると、概念語の数を N とし、表 2 のようになる。

4. 評価

本章では、多様な手がかりを同時に考慮することで、より良い順序が得られることを確かめるため、大規模なソーシャルメディアテキストを解析対象とし、客観的・主観的評価に基づくクエリについて既存手法と提案手法の比較を行う。

4.1 データ

4.1.1 解析対象テキスト

順序付けのための手がかりを抽出する知識源として、本稿では、我々が収集した 200 万記事に及び 2005 年から 2013 年の日本語のブログ記事と、2011 年から 2013 年の Twitter への投稿を利用した。ブログ記事は、100 万人を超えるユーザによって書かれた約 20 億の文であり、Twitter への投稿は、290 万人近くのユーザによって投稿された約 60 億文を利用した。各テキストは、Kaji ら [8] の手法を用いて形態素解析を行い、Yoshinaga ら [18] が開発した J.DepP^(注3) により、係り受け解析を行っている。

(注2): PMI の計算においては、対数をとる際に、頻度に 1 を加算するスムージングによって 0 頻度のものにも対応している。

(注3): <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/jdepp/>

表 3 形容詞と概念語集合の組み合わせ

(a) 客観的評価に基づくクエリ

概念語カテゴリ	形容詞	概念語集合 (調査に基づく正解 . 左ほど性質が強い .)
動物	大きい	クジラ, ソウ, ウマ, イヌ, ネコ, ネズミ
乗り物	速い	飛行機, 新幹線, 電車, タクシー, バス, 自転車
プログラミング言語	遅い	Ruby, Python, Perl, Java, JavaScript, Lisp, Scala, Haskell
果物	小さい	さくらんぼ, イチゴ, あんず, みかん, モモ, りんご, メロン
国	温暖だ	インド, タイ, スペイン, イギリス, ロシア
お寺 (京都)	古い	法隆寺, 善光寺, 薬師寺, 東寺, 中尊寺, 増上寺, 金閣寺, 銀閣寺
邦画	長い	ROOKIES, GOEMON, ソラニン, フラガール, サマーウォーズ, 火垂るの墓, 羅生門
芸能人 (女性)	若い	北乃きい, 南明奈, 優香, 吹石一恵, 釈由美子, 友近, 梨花
電気メーカー	古い	NEC, 東芝, 日立製作所, 富士通, パナソニック, キヤノン, エプソン, ソニー
コンビニ	多い	セブンイレブン, ローソン, ファミマ, ミニストップ, セイコーマート

(b) 主観的評価に基づくクエリ

概念語カテゴリ	形容詞	概念語集合 (人手による順序付けに基づく正解 . 左ほど性質が強い .)
花	綺麗だ	サクラ, バラ, 百合, ラベンダー, 桔梗, ひまわり, 椿, 菊
宝石	上品だ	サファイア, エメラルド, パール, ルビー, アメジスト, オパール, トルマリン, ターコイズ
アルコール	美味しい	ビール, ワイン, シャンパン, 焼酎, チューハイ, ハイボール, テキーラ, マッコリ
スポーツ	楽しい	サッカー, 卓球, バスケ, 相撲, テニス, バレー, 野球, プロレス
動物	賢い	イヌ, クジラ, ネコ, ソウ, ネズミ, ウマ
乗り物	快適だ	新幹線, タクシー, 飛行機, 自転車, バス, 電車
食べ物	美味しい	ステーキ, ラーメン, パスタ, カレー, ピザ, チャーハン, ハンバーガー
楽器	心地よい	フルート, チェロ, クラリネット, オルガン, トランペット, ギター, ハーモニカ, ドラム
プログラミング言語	簡単だ	Ruby, Python, Perl, Java, JavaScript, Lisp, Scala, Haskell
動物	かわいい	リス, ウサギ, イヌ, ペンギン, パンダ, ウマ, トカゲ, ライオン
野菜	美味しい	ほうれん草, 玉ねぎ, かぼちゃ, ナス, ブロッコリー, 白菜, きゅうり, もやし
果物	甘い	メロン, モモ, りんご, さくらんぼ, イチゴ, みかん, あんず
電化製品	便利だ	スマートフォン, パソコン, デジカメ, カーナビ, プリンタ, カメラ, スピーカー
食用肉	好きだ	鶏肉, 牛肉, 豚肉, ラム肉, 猪肉, 鹿肉, 馬肉
鳥類	かわいい	ペンギン, ふくろう, うずら, スズメ, 白鳥, にわとり, キジ, 鷺
天候	嫌いだ	黄砂, 雨, 雷, 強風, 霧, 雪, 霜, 晴天
国	安全だ	イギリス, タイ, スペイン, インド, ロシア
お寺 (京都)	有名だ	金閣寺, 銀閣寺, 法隆寺, 薬師寺, 善光寺, 中尊寺, 東寺, 増上寺
歌手 (日本)	好きだ	ボルノグラフィティ, スピッツ, TOKIO, スキマスイッチ, SMAP, コブクロ, ZARD, EXILE
アニメ	面白い	ガンダム, ドラゴンボール, ONEPIECE, バガボンド, こち亀, ガッチャマン, ヤッターマン, 美味しんぼ
邦画	面白い	サマーウォーズ, フラガール, 羅生門, ソラニン, 火垂るの墓, ROOKIES, GOEMON
芸能人 (女性)	可愛い	釈由美子, 北乃きい, 吹石一恵, 優香, 梨花, 南明奈, 友近
電気メーカー	有名だ	ソニー, パナソニック, 東芝, NEC, 日立製作所, 富士通, キヤノン, エプソン
野球チーム (メジャー)	有名だ	ヤンキース, マリナーズ, レッドソックス, ドジャース, メッツ, ホワイトソックス, オリオールズ, インディアンズ
ファストフード	美味しい	モスバーガー, フレッシュネスバーガー, ケンタッキー, ミスタードーナツ, バーガーキング, マクドナルド
車メーカー	好調だ	トヨタ, ホンダ, ヤマハ, マツダ, ダイハツ
コンビニ	便利だ	セブンイレブン, ローソン, ファミマ, セイコーマート, ミニストップ
ブラウザ	快適だ	Chrome, FireFox, Safari, Opera, Sleipnir
都市名	安全だ	ハワイ, ロンドン, ベルリン, パリ, 香港, シカゴ, ローマ, モスクワ
カフェ	好きだ	スターバックス, サンマルク, タリーズ, プロント, ドトール, エクセルシオール, ルノアール
地名 (東京)	おしゃれだ	青山, 渋谷, 新宿, 品川, 中野, 池袋, 上野, 浅草
出版社	有名だ	集英社, 岩波書店, 新潮社, 幻冬舎, 白泉社, 祥伝社, 東京創元社

4.1.2 クエリ

実験のためのクエリ (概念語集合と形容詞の組) はブログ記事を次のように解析することで集めた。まず、ブログ記事の中で最も記事数が多い 2009 年のデータから 10 分の 1 の記事をサンプリングし、そこに含まれる単語を Brown クラスタリング [3] によってクラスタリングした。次に、それらのクラスタを調べて、共通のカテゴリにある最大 8 個の概念語のグループを作った。最後に、これらの概念語グループについて、それぞれの概念語との自己相互情報量 PMI の平均をもとに、共通す

る性質の形容詞を決め、概念語集合と形容詞の組を作成した。

クエリは、「お寺」の「古い」順 (建立されてからどのくらい経っているか) などの数値で表せる特徴に基づくものを「客観的評価に基づくクエリ」、それ以外を「主観的評価に基づくクエリ」として区別した。得られたクエリの一覧は表 3 に示す。それぞれのクエリの正解は以下のように作成した。

客観的評価に基づくクエリ 客観的評価に基づくクエリでは、形容詞が表す性質が数値で表せる特徴であるため、それぞれの形容詞に基いて概念語を調査し、その結果から正解の順序付け

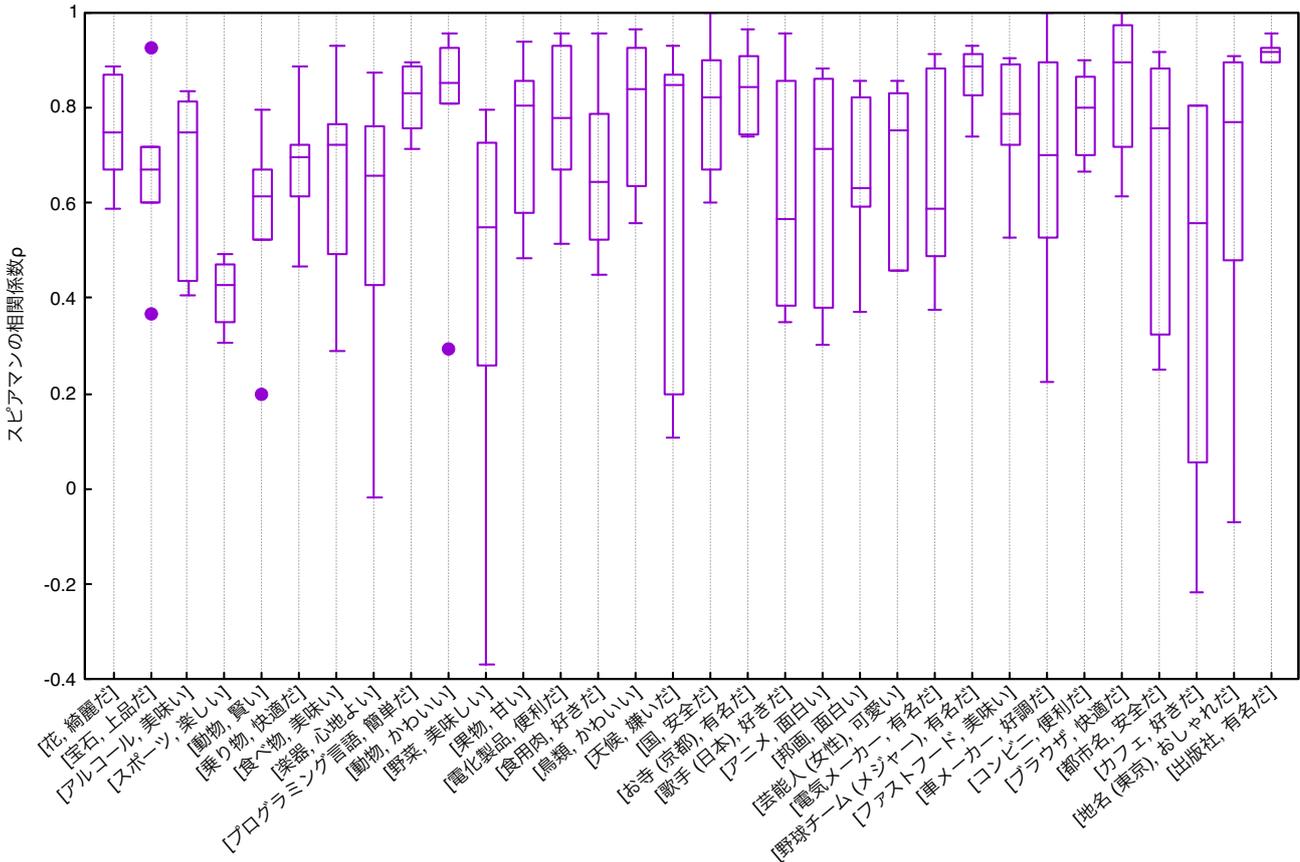


図1 人手で作成した順序と正解との相関係数

を与えた。このようにして得られた正解を表 3(a) に示す。主観的評価に基づくクエリ 主観的な評価に基づくクエリでは、コンピュータ科学に関わりのある、20 代から 30 代の学生 (3 人) や教員 (3 人), システムエンジニア (1 人) の合計 7 人 (男性 5 人, 女性 2 人) の被験者にそれらの概念語の集合と形容詞の組を提示し、順序付けを行ってもらった。この 7 つの人手による順序付けと概念語の考えられるすべての順序との、スピアマンの順位相関係数 ρ を計算し、その平均が最も大きくなる順序を正解とした。得られた正解は表 3(b) に示す。また、人手で作成した順序と正解の順序との相関係数に関する箱ひげ図を図 1 に示す。

主観的評価に基づくクエリについて、図 1 を見ると、全体としてはある程度の相関が得られているが、楽器 [心地よい] や野菜 [美味しい], カフェ [好きだ] などの結果は、個人によって非常に差がある一方、コンビニ [便利だ] や出版社 [有名だ] などは、ばらつきが小さいことがわかる。また、宝石 [上品だ] や動物 [かわいい] などは、ばらつきが小さいものの外れ値となる回答も見られた。このことは、主観的評価に基づくクエリを解くことが本質的に難しいことを表している。

4.2 学習とテスト

学習とテストにおいては、以上のような客観的評価に基づくクエリと主観的評価に基づくクエリをそれぞれ別々の問題として取り扱う。また、学習に用いる特徴量としては、以下の 3 種類を用いた。

- (1) ブログ記事から得た特徴量 \vec{v}_b
- (2) ツイートから得た特徴量 \vec{v}_t
- (3) (1) と (2) を合わせた特徴量 (\vec{v}_b, \vec{v}_t)

よって、合計 6 通りのクエリとテキストデータの組み合わせについて、学習とテストを行った。

テストにおいては、表 3 に示す客観的評価に基づくクエリ、主観的評価に基づくクエリそれぞれについて、概念語の集合と形容詞の組から 1 つをテストデータとし、他のデータを教師データとする leave-one-out 交差検定を行い、提案手法から得られた順序付けと、正解の順序とのスピアマンの順位相関係数 ρ を計算した。また、ランキング SVM, SVR には LIBLINEAR (注4) を用い、学習には線形カーネルを利用した。また、各テストにおけるランキング SVM, SVR のハイパーパラメータ C は教師データの中で交差検定を行って最適な値を求めた。

4.3 結果

結果は、表 4 のとおりであり、解析対象のテキストの種類ごとに、最も相関係数が高い手法に下線を引き、すべての手法の中で最も良い相関係数のものには、二重下線を引いている。ベースラインとしては、仁科ら [20] の手法において、最も性能が良い、係り受けを手がかりとする手法を用いている。結果として、客観的評価に基づくクエリ、主観的評価に基づくクエリともに、いずれのソーシャルメディアテキストを用いた場合も、

(注4): <https://www.csie.ntu.edu.tw/~cjlin/liblinear/>

表 4 実験結果

(a) 客観的評価に基づくクエリの結果

概念語 [形容詞]	Blog			Tweet			Blog + Tweet	
	BASELINE	SVM	SVR	BASELINE	SVM	SVR	SVM	SVR
動物 [大きい]	<u>0.943</u>	0.886	<u>0.943</u>	0.600	<u>0.829</u>	0.600	0.771	<u>0.886</u>
乗り物 [速い]	0.257	<u>0.600</u>	0.257	0.257	0.543	<u>0.600</u>	<u>0.600</u>	<u>0.600</u>
プログラミング言語 [遅い]	-0.024	0.071	<u>0.119</u>	-0.024	0.190	<u>0.500</u>	0.048	<u>0.619</u>
果物 [小さい]	0.286	<u>0.536</u>	0.321	0.464	0.536	<u>0.607</u>	<u>0.393</u>	0.321
国 [温暖だ]	<u>0.900</u>	0.600	0.700	0.700	<u>0.900</u>	<u>0.900</u>	0.700	<u>0.900</u>
お寺 (京都) [古い]	0.500	<u>0.595</u>	0.571	<u>0.976</u>	0.952	0.952	<u>0.762</u>	0.595
邦画 [長い]	0.321	<u>0.357</u>	0.250	-0.071	<u>0.429</u>	<u>0.429</u>	0.321	<u>0.571</u>
芸能人 (女性) [若い]	-0.179	<u>0.321</u>	0.179	0.143	0.143	0.143	0.071	<u>0.321</u>
電気メーカー [古い]	0.190	<u>0.310</u>	0.262	0.738	0.429	<u>0.786</u>	0.262	<u>0.452</u>
コンビニ [多い]	<u>0.300</u>	<u>0.300</u>	<u>0.300</u>	<u>-0.100</u>	<u>-0.100</u>	-0.600	-0.100	-0.100
平均	0.350	<u>0.458</u>	0.390	0.368	0.485	<u>0.492</u>	0.383	<u>0.517</u>

(b) 主観的評価に基づくクエリの結果 (正解データと人手による順序付けとの相関係数の平均を HUMAN として参考を示した)

概念語 [形容詞]	HUMAN	Blog			Tweet			Blog + Tweet	
		BASELINE	SVM	SVR	BASELINE	SVM	SVR	SVM	SVR
花 [綺麗だ]	0.749	<u>0.286</u>	0.167	0.095	0.190	<u>0.476</u>	<u>0.476</u>	<u>0.238</u>	0.214
宝石 [上品だ]	0.667	0.238	<u>0.476</u>	0.357	0.214	0.381	<u>0.524</u>	0.619	<u>0.857</u>
アルコール [美味しい]	0.648	0.167	0.690	<u>0.857</u>	0.524	<u>0.762</u>	<u>0.762</u>	0.667	<u>0.786</u>
スポーツ [楽しい]	0.412	0.238	0.310	<u>0.381</u>	<u>0.333</u>	0.286	0.024	<u>0.524</u>	0.333
動物 [賢い]	0.578	-0.200	0.143	<u>0.257</u>	<u>0.600</u>	0.086	0.257	0.143	0.143
乗り物 [快適だ]	0.683	<u>0.371</u>	0.257	0.257	<u>0.486</u>	<u>0.486</u>	<u>0.486</u>	<u>0.486</u>	0.257
食べ物 [美味しい]	0.639	0.143	0.393	<u>0.679</u>	0.143	<u>0.500</u>	0.286	<u>0.393</u>	0.214
楽器 [心地よい]	0.570	-0.048	0.095	<u>0.119</u>	-0.595	<u>-0.333</u>	-0.357	<u>-0.190</u>	-0.214
プログラミング言語 [簡単だ]	0.826	0.476	0.619	<u>0.786</u>	0.762	0.881	<u>0.905</u>	0.786	0.810
動物 [かわいい]	0.790	<u>0.738</u>	0.571	0.667	0.214	0.500	<u>0.643</u>	0.524	<u>0.738</u>
野菜 [美味しい]	0.451	<u>0.524</u>	0.429	0.262	<u>0.071</u>	-0.286	-0.048	-0.429	-0.024
果物 [甘い]	0.729	<u>0.964</u>	0.607	0.643	<u>0.857</u>	0.607	0.643	<u>0.750</u>	0.643
電化製品 [便利だ]	0.772	0.536	<u>0.679</u>	<u>0.679</u>	0.143	0.750	<u>0.857</u>	0.714	<u>0.786</u>
食用肉 [好きだ]	0.662	-0.429	<u>0.179</u>	0.107	-0.607	<u>-0.286</u>	-0.607	0.000	<u>0.500</u>
鳥類 [かわいい]	0.800	0.881	<u>0.929</u>	0.905	<u>0.929</u>	0.810	<u>0.929</u>	0.905	0.905
天候 [嫌いだ]	0.651	0.738	0.690	<u>0.833</u>	0.810	<u>0.857</u>	0.595	<u>0.738</u>	0.595
国 [安全だ]	0.804	-0.500	-0.200	<u>0.000</u>	<u>-0.300</u>	-0.600	<u>-0.300</u>	-0.700	<u>0.200</u>
お寺 (京都) [有名だ]	0.841	0.190	0.643	<u>0.762</u>	0.524	0.429	<u>0.643</u>	<u>0.762</u>	0.690
歌手 (日本) [好きだ]	0.614	0.762	0.667	<u>0.857</u>	<u>0.857</u>	0.571	0.095	<u>0.548</u>	0.048
アニメ [面白い]	0.633	-0.167	<u>0.429</u>	-0.238	<u>0.738</u>	<u>0.738</u>	<u>0.738</u>	0.524	<u>0.571</u>
邦画 [面白い]	0.649	-0.071	-0.107	<u>-0.036</u>	0.107	<u>0.286</u>	0.107	0.000	<u>0.286</u>
芸能人 (女性) [可愛い]	0.699	0.000	<u>0.071</u>	0.000	0.250	<u>0.286</u>	0.214	<u>0.429</u>	0.214
電気メーカー [有名だ]	0.644	0.381	<u>0.643</u>	0.333	<u>0.810</u>	0.738	0.643	<u>0.857</u>	0.690
野球チーム (メジャー) [有名だ]	0.864	<u>0.976</u>	0.905	0.952	0.762	0.762	<u>0.810</u>	<u>0.929</u>	<u>0.952</u>
ファストフード [美味しい]	0.774	<u>0.486</u>	-0.086	-0.371	<u>0.771</u>	0.486	-0.086	<u>0.086</u>	-0.029
車メーカー [好調だ]	0.665	-0.900	-0.700	<u>0.000</u>	-0.100	0.400	<u>0.900</u>	<u>0.700</u>	0.000
コンビニ [便利だ]	0.791	0.400	<u>0.600</u>	0.300	0.100	<u>0.400</u>	0.100	<u>0.500</u>	<u>0.500</u>
ブラウザ [快適だ]	0.856	-0.200	<u>-0.100</u>	<u>-0.100</u>	<u>0.700</u>	<u>0.700</u>	<u>0.700</u>	-0.100	<u>0.500</u>
都市名 [安全だ]	0.649	0.429	<u>0.762</u>	0.214	-0.071	<u>0.190</u>	0.143	0.048	<u>0.310</u>
カフェ [好きだ]	0.405	0.071	<u>0.571</u>	0.393	<u>0.357</u>	0.179	0.179	<u>0.607</u>	0.286
地名 (東京) [おしゃれだ]	0.658	0.524	0.595	<u>0.738</u>	0.571	0.619	<u>0.714</u>	<u>0.571</u>	0.333
出版社 [有名だ]	0.916	0.786	<u>0.964</u>	0.857	0.893	<u>0.929</u>	0.893	<u>0.929</u>	0.893
平均	0.690	0.275	<u>0.403</u>	0.392	0.376	<u>0.425</u>	0.402	0.424	<u>0.437</u>

提案手法のランキング SVM, SVR の作成した順序と, 正解の順序との相関係数の平均がベースラインを上回る結果となって

おり, 教師ありランキング学習によって様々な手がかりを組み合わせた効果が見られる.

特に、ブログ記事とツイートの両方を用いた SVR による順序が客観的評価に基づくクエリ、主観的評価に基づくクエリのいずれでも相関係数の平均が最も高くなり、それぞれ 0.517、0.437 という結果となった。

また、主観的評価に基づくクエリでは、半数以上の 32 個中 18 個において、提案手法の少なくとも 1 つが、人手による順序と正解の順序との平均相関係数を上回る結果となっている。この結果から、絶対的正解のある客観的評価に基づくクエリだけでなく、人の最大公約数的意見を反映した主観的評価に基づくクエリについても提案手法は有効であるといえる。

それぞれの結果をより詳しく見ると、まず客観的評価に基づくクエリでは、ブログ記事を用いた場合、提案手法ではすべての場合において正の相関があり、提案手法が非常に有用であることがわかる。一方で、ツイートを用いた場合は、コンビニ [多い] の結果に負の相関が見られるが、全体の相関係数の平均はブログ記事よりも良くなっていることが見て取れる。最後に、ブログ記事とツイートの両方の特徴量を組み合わせた場合は、ランキング SVM の結果が悪くなっている一方で、SVR の結果は向上し、SVR のほうがより多くの特徴を用いる効果が大きいことがわかる。

このことは、主観的評価に基づくクエリにおいても同様で、ブログ記事とツイートの両方の特徴量を組み合わせた場合は、ランキング SVM ではそれらを独立して用いた場合とほとんど変わらないが、SVR では相関係数の平均が大きくなっており、最も良い結果となっている。また、ツイートを用いた結果では、客観的評価に基づくクエリとは異なり、相関係数の平均にブログ記事を用いた場合からの大きな向上は見られない。

さらに、全体を見ると、ツイートを用いた結果のほうがブログ記事を用いた結果よりも分散が大きくなっていることがわかる。

4.4 考察

これらの結果について、それぞれの手がかりがどのように影響しているかを調査するために、主観的評価に基づくクエリに対してブログ記事を用いた、いくつかの順序付けを確認した。文内共起 電気メーカー [有名だ] については、ソーシャルメディアテキストにおいて、概念語と「有名だ」との間の共起がよく現れ、これによってソニー、パナソニック、東芝の順位が上昇し、正解との相関が高くなっている。これは企業について有名であるということを手がかりとすることが多くなく、有名な商品などと一緒に共起することが間接的に手がかりとなったからと考えられる。

係り受け 果物 [甘い] において、正解の順序と同様に概念語と形容詞の係り受けの関係が成り立っていることがベースラインを見るとわかり、結果としてこの手がかりも組み合わせて利用している提案手法でも相関が高くなっている。

直喩表現 鳥類 [かわいい] では、直喩表現がペンギンについて多く現れ、これがランキング SVM、SVR においてペンギンの順位を 1 位とすることで正解との相関を押し上げている。

比較表現 比較表現が相関の向上に寄与している例としては、カフェ [好きだ] のスターバックスがあり、ランキング SVM、

SVR ともに順位をあげている。

以上のように様々な手がかりを組み合わせることが、正解により近い順序を作成することにつながっていると考えられる。

一方で、ベースライン、提案手法ともに相関係数が低くなる、主観的評価に基づくクエリ、楽器 [心地良い] は、図 1 で見たように個人のばらつきが大きい問題であり、集めた手がかりでは正しく学習が行えなかったと推察できる。同様のことが、野菜 [美味しい] をツイートを用いて解いた結果にも言える。これに対して、人によるばらつきが小さく、相関係数が高くなっている例として、主観的評価に基づく動物 [かわいい] があるが、これは直接的に事物についての特徴を表す形容詞「かわいい」の言及される回数が多いことによって、十分な手がかりを集められたと推察できる。

また、主観的評価に基づく動物 [賢い] について、提案手法の相関係数が低くなっているのは、意外性のあること (例: 「ネズミって思ったよりも賢いな」) や限定的な個体を指す表現 (例: 「近所の猫は賢い」) などにより、共起や係り受けの回数が増加し、それぞれの概念語の相対的な性質の強さとは異なる結果となったことが原因の 1 つとして考えられる。

ところで、客観的評価に基づくクエリにおいて、プログラミング言語 [遅い] やコンビニ [多い] は比較的相関係数が低くなっている。これらは時間的に変化する特徴についての順序付けであり、手がかりを集めたソーシャルメディアテキストが数年にわたるものであったため、これらの時間による変化が順序付けに影響したと考えられる。

5. まとめと今後の課題

本研究では、与えられた概念語の集合 (例: ロンドン、ハワイ、ローマ) を、形容詞 (例: 安全だ) が表す性質の強さによって順序付けするタスクに取り組んだ。提案手法では、大規模なソーシャルメディアテキストを解析し、既存手法で集めていた、形容詞と概念語の文内共起や係り受けといった手がかりだけでなく、直喩表現や比較表現も数えあげ、ランキング SVM と SVR による教師ありランキング学習を用いることでそれらの手がかりを組み合わせる順序付けを行った。評価では、実際のブログ記事やツイートを解析対象として手法を適用し、客観的評価に基づくクエリでは調査に基づく正解の順序、主観的評価に基づくクエリでは人手によって作成した正解の順序を、提案手法が選んだ順序と比較することで実験を行った。結果として、いずれのデータセットについても、提案手法が既存手法を上回り、手法の有用性が確かめられた。

本研究では、ランキング SVM や SVR といった教師ありランキング学習を用いて既存手法では独立に用いられていた多様な手がかりを、組み合わせることを可能とした。このことにより、手法を改善する際に、新しい手がかりを加えることを容易にすると同時に、本稿で行ったようにブログ記事とツイートから得た特徴量を組み合わせるなどの工夫ができるようになったといえる。

本研究では、既存手法に比べて、提案手法が多くの場合に優れていることが示されたが、限定的な条件や意外性のあるこ

とを明言するような場合において、相関係数が低くなる場合があった。そのため、それらの文を手がかりに加えることが改善方法の1つとして考えられる。例えば、「この○○(概念語)は(形容詞)」などのような概念語を限定するパターンを考慮することで、精度の向上がはかれると考えられる。

また、現在の手法では、与えられた概念語と形容詞のみの関係を数え上げるため、それらの共起頻度が大きくないとそもそも順序付けが難しいという問題がある。これに対して、概念語の上位語・下位語を検索の対象に含める(例:「車」に対して「タクシー」も考える)などの工夫ができると考えられる。また、形容詞についても同様のことが言え、与えられた形容詞(あるいはその対義語)の類義語や、関連語(例:「大きい」ものは「重い」ことが多い)を前処理で作成し、これらについても数え上げることでより多くの手がかりを用いることができる。

これらの改善を取り入れた上で、より多くのクエリについて取り組むことも重要であると考えられる。そのために、クラウドソーシングなどを用いて、評価を行うことが考えられる。この際には、順序付けする観点として、形容詞で表現された概念語の属性に限らず、形容表現(例:動物の概念語に対して「しっぽが長い」)に関する順序付けも行うことで、取り扱える問題を一般化できると考えられる。

さらに、本稿では、日本語のブログ記事やツイートから、それらの利用者全体の集合知としての順序付けを得ることを試みたが、手法を適用するテキストを絞って、ある集団に特有の順序付け(例:南米の人の食べ物の好み)を見つけるなどの応用が可能か検証することも興味深い。また、実験の考察で時間に依って変化する特徴に関する順序付けが難しいことを述べたが、解析対象の別の絞り方として、年ごとに手がかりを集めて順序付けを行うことで、例えば政党の支持率の推移を見るなどの応用も考えられる。

謝 辞

本研究の一部は JSPS 科研費 25280111 の助成を受けたものです。

文 献

- [1] Sören Auer and Jens Lehmann. What Have Innsbruck and Leipzig in Common? Extracting Semantics from Wiki Content. In *Proceedings of the 4th European Conference on The Semantic Web (ESWC)*, pp. 503–517, 2007.
- [2] Ralph A. Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. Vol. 39, pp. 324–345, December 1952.
- [3] Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, and Jennifer C. Lai. Class-Based n -gram Models of Natural Language. *Computational Linguistics*, Vol. 18, No. 4, pp. 467–479, December 1992.
- [4] Hsin-Hsi Chen, Shih-Chung Tsai, and Jin-He Tsai. Mining tables from large scale HTML texts. In *Proceedings of the 18th conference on Computational linguistics (COLING)*, pp. 166–172, 2000.
- [5] Xi Chen, Paul N. Bennett, Kevyn Collins-Thompson, and Eric Horvitz. Pairwise ranking aggregation in a crowd-sourced setting. In *Proceedings of the Sixth ACM In-*

ternational Conference on Web Search and Data Mining (WSDM), pp. 193–202, 2013.

- [6] Harris Drucker, Christopher J. C. Burges, Linda Kaufman, Alexander J. Smola, and Vladimir N. Vapnik. Support Vector Regression Machines. In *Advances in Neural Information Processing Systems 9, NIPS 1996*, pp. 155–161. MIT Press, 1997.
- [7] Thorsten Joachims. Optimizing Search Engines Using Clickthrough Data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 133–142, 2002.
- [8] Nobuhiro Kaji and Masaru Kitsuregawa. Efficient Word Lattice Generation for Joint Word Segmentation and POS Tagging in Japanese. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 153–161, Nagoya, Japan, October 2013. Asian Federation of Natural Language Processing.
- [9] Takeshi Kurashima, Katsuji Bessho, Hiroyuki Toda, Toshio Uchiyama, and Ryoji Kataoka. Ranking Entities Using Comparative Relations. In *Proceedings of the 19th Conference on Database and Expert Systems Applications (DEXA)*, pp. 124–133, 2008.
- [10] Shuzi Niu, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. Stochastic rank aggregation. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 478–487, 2013.
- [11] Bo Pang and Lillian Lee. *Opinion Mining and Sentiment Analysis*. Now Publishers Inc., 2008.
- [12] John Prager. *Open-Domain Question Answering*. Now Publishers Inc., 2007.
- [13] Karthik Raman and Thorsten Joachims. Methods for ordinal peer grading. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 1037–1046, 2014.
- [14] Hiroya Takamura and Jun'ichi Tsujii. Estimating Numerical Attributes by Bringing Together Fragmentary Clues. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1305–1310, 2015.
- [15] Peter Turney. Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 417–424, 2002.
- [16] Maksims N. Volkovs and Richard S. Zemel. A flexible generative model for preference aggregation. In *Proceedings of the 21st International Conference on World Wide Web (WWW)*, pp. 479–488, 2012.
- [17] Fei Wu and Daniel S. Weld. Autonomously semantifying Wikipedia. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (CIKM)*, pp. 41–50, 2007.
- [18] Naoki Yoshinaga and Masaru Kitsuregawa. Kernel Slicing: Scalable Online Training with Conjunctive Features. In *COLING*, pp. 1245–1253. Tsinghua University Press, 2010.
- [19] Naoki Yoshinaga and Kentaro Torisawa. Open-domain attribute-value acquisition from semi-structured texts. In *Proceedings of the 6th International Semantic Web Conference (ISWC-07), Workshop on Text to Knowledge: The Lexicon/Ontology Interface (OntoLex-2007)*, pp. 55–66, 2007.
- [20] 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. *IPSJ SIG NL*, Vol. 2013, No. 8, pp. 1–7, nov 2013.