

tweet を用いた階層的同位語の感性的結びつきの時間的可視化

中井戸 晃彦^a 山西 良典^b 福本 淳一^b

立命館大学情報理工学部メディア情報学科

a) a_nakaido@nlp.is.ritsumei.ac.jp b) {[ryama](mailto:ryama@media.ritsumei.ac.jp), [fukumoto](mailto:fukumoto@media.ritsumei.ac.jp)}@media.ritsumei.ac.jp

概要 本稿では、tweet 中の情報を利用し、時間的に変化する階層的同位語の感性的結びつきを可視化する。階層的同位語とは、階層知識の中で共通の上位概念を持つ知識のことである。人はある知識について説明する場合、階層的同位語の中で対象知識に対して多くの人が共通して抱く印象が類似している知識を用いることで、対象知識の印象などの感性情報を伝える。知識についての感性情報は時間的に変化するため、印象が類似した知識を獲得するためには、静的な情報だけでなく、動的に更新される情報を参照すべきである。提案手法では、時間的に変化する情報が含まれる tweet 中の情報を利用することにより、時間的に変化する階層的同位語の感性的結びつきを算出する。そして、対象知識と感性的に強く結びつく階層的に同位な知識を感性的に比較可能な知識とし、対象知識に対して感性的に比較可能な知識が時間的に変化する様子を可視化する。

キーワード 感性, Twitter, 時間的可視化, 階層知識

1 はじめに

人はある知識について説明する場合、辞書などを参照して関連する情報を列挙することで、対象知識の詳細を伝える。しかしながら、ある知識についての印象などの感性情報は時間的に変化するため、辞書などの静的な情報のみを用いて説明することは難しい。このような場合、人間同士のコミュニケーションでは、感性的に比較可能な他の知識を用いることで、対象知識の印象を伝えている。例えば“新垣結衣”を説明する場合、新垣結衣と感性的に比較可能な日本の女優である“堀北真希”や“上戸彩”を挙げることで、これらの他の語句に紐ついた情報から“新垣結衣”の特徴や印象を連想することができる。

ある知識に対して比較可能な知識として、階層知識の中で共通の上位概念を持つ知識（階層的同位語）が用いられる。階層的同位語の中でも、対象知識に対して印象が類似した知識が、感性的に比較可能と言える。このとき、知識に対する多くの人が共通して抱く印象は時間的に変化すると思われる。例えば、人は“嵐山”に対して、春には「桜の名所」という印象を抱く一方で、秋には「紅葉の名所」という印象を抱く。この例では、春には桜の名所、秋には紅葉の名所を“嵐山”の感性的に比較可能な他の知識としてそれぞれ挙げるべきである。したがって、感性的に比較可能な他の知識を獲得するためには、辞書などの静的な情報だけでなく、動的に更新される情報を参照すべきであると考えられる。

提案手法では、時間的に変化する情報が含まれる Twitter の投稿である tweet 中の情報を利用し、時間的に変化する階層的同位語の感性的結びつきを算出する。そし

表1 “タレント” “日本の女優” “歌手” を上位概念に持つ階層的同位語の例。

上位語	下位語
“タレント”	“中川翔子” “大島優子” “矢口真里” “外国人タレント” “子役タレント” and etc.
“日本の女優”	“戸田 恵梨香” “堀北真希” “大島優子” “その他の女優” “上戸彩” and etc.
“歌手”	“桑田佳祐” “大島優子” “松任谷由実” “西野カナ” “安室奈美恵” and etc.

て、階層的同位語の中でも、対象知識と感性的に強く結びつく知識を感性的に比較可能な知識とし、対象知識に対して感性的に比較可能な知識が時間的に変化する様子を可視化する。また本稿では、感性的結びつきに寄与した情報を考察することで、対象知識と階層的同位語の感性的結びつきの様子が変化した要因について議論する。

2 階層知識中の比較可能な知識

知識体系は、一般的に上位下位の概念構造によって説明できる。例えば、“ピアノ”の上位語は“楽器”であり、逆に“ピアノ”は“楽器”の下位語である。階層知

識は、最も一般的な概念を最上位として、順により詳細な分類を示す概念からなる階層構造を成している。階層知識の代表的なものとしては、クラスライブラリやインターネットディレクトリ、図書の目録分類、オントロジなどが挙げられる。知識間の階層概念を構築する手法についても、数多く報告されている [1, 2]。

階層知識の中で共通の上位概念を持つ知識は、階層的に同位であり、任意の対象知識に対しての比較可能な知識と成り得る。表 2 に、“タレント”“日本の女優”“歌手”の下位語の例を示す。同表より、“大島優子”“中川翔子”“矢口真里”は、“タレント”という共通の上位概念を持つ階層的な同位語であることがわかる。また、“大島優子”は“タレント”だけでなく、“日本の女優”“歌手”などの他の上位概念を持っており、複数の上位概念を持つ知識といえる。実社会のコミュニケーションでは、「大島優子のようなタレント」や「大島優子のような日本の女優」というように上位概念を指定することで、どの上位概念における同位語であるかを特定している。

階層的な同位語は、広義には対象知識として比較可能であるが、不適当なものが含まれている場合がある。例えば、“タレント”という上位概念を持つ“外国人タレント”“子役タレント”や、“日本の女優”という上位概念を持つ“その他の女優”は、階層関係として適当である。しかしながら、一般的なコミュニケーションにおいて“大島優子”の比較対象として、“外国人タレント”“子役タレント”、“その他の女優”などの概念的な知識を比較対象として挙げることはなく、直感的に不適当な比較対象と考えられる。また、“大島優子”は女性アイドルグループである“AKB48”に所属しており、“大島優子”の歌手としての評価は“AKB48”というグループ単位で評価されることが一般的である。一方で、“大島優子”と共通の上位概念“歌手”を持つ“松任谷由実”や“西野カナ”は、ソロで活躍している歌手であるため、“松任谷由実”“西野カナ”としてそれぞれ評価される。そのため、これらの知識については歌手の“大島優子”の感性的な比較対象としては不適当と考える。

これらの問題を解決し、直感的に適切な同位語を取得するためには、階層知識の構築時に参考される広義な情報ではなく、印象や評判などの感性的な情報を取り入れる必要がある。このとき、知識に関する印象や評判、紐づく関連情報は時間的に変化していくため、ダイナミック性に富んだ感性的な情報を取得することで、より直感的に適切な同位語の取得が可能になると考える。

3 対象知識に関する感性的な情報の獲得

本節では、前節で述べた階層知識の問題点を解決するために利用する情報の知識源について考察する。また、tweet 中の時間的に変化する情報と感性的な意見の投稿について詳細に議論する。

3.1 対象知識に紐づく情報の知識源

人は知識に紐づく情報から、知識の特徴や印象を連想する。山西らは、知識に紐づく情報を獲得する情報源として、Web スニペットを利用している [3]。山西らの手法では、ページの内容からクエリが出現している部分の周辺を抜粋して表示する Web スニペットを用いて、対象知識に紐づく情報を獲得している。Web スニペットの利用は、公式サイトやニュースサイトなどに含まれる身長、体重などの数値データや職業、組織などの客観的な情報については比較的取得が容易と考える。一方で、「可愛い」「楽しい」「美味しい」などの対象知識についての特徴や印象を表す評判情報については、取得が難しいと考える。

Web スニペットに比べて、より主観的な意見が多いと考えられる媒体にブログがある。ブログでは、作者の個人的な体験や日記、特定のトピックに関する話題などを Web 上に記録・公開するため、感想や主観的な意見が多い傾向にある。そのため、ブログを対象知識の情報源として用いることで、客観的な情報だけでなく、評判情報についてもより多く取得可能な情報源であると考えられる。しかしながら、ブログは体験や話題の内容を文章形式でまとめて記述し投稿されるため、日々刻々と変化する感性的な評判を捉えるためのダイナミック性を持ち合わせているとは言えない。

よりダイナミック性が高く、主観的な評判情報が多いと考えられる知識源として Twitter が考えられる。Twitter では、任意の対象についての意見や感想が 140 文字の制限の中で投稿される。Twitter の投稿 (tweet) 中には主観的な意見が多く感性的な評価を取得するうえで有用であると同時にダイナミック性にも優れており、時間的に変化する知識に対する印象を取得するための知識源として有用性が高いと考える。Twitter 上の情報を用いることで、世論の直感的な感想の取得 [4] やユーザが着目するトピックの分析の研究なども報告されている [5]。

3.2 Twitter のダイナミック性

Twitter は、世界で 2 億人以上のユーザが利用している代表的なマイクロブログの 1 つであり、ユーザ間で意見共有や感想共有のツールとして広く知られている。Twitter はパソコンだけでなく、スマートフォンやタブレット PC などの携帯端末からも投稿可能であり、多くのユーザが気軽に利用可能である。そのため、現在の行動や意見、感想が気軽に投稿され、ダイナミックに情報

- HIRA(@RedBlaze184)
“ TBS に大島優子さん出てる。 ”
2013 年 11 月 5 日,14:16.Tweet.
- かねっち (@yra503_)
“ 今日も大島優子可愛いなー !!! ”
2013 年 11 月 4 日,21:15.Tweet.
- すずりょー (@sizr49)
“ 安堂ロイド面白い！大島優子と柴咲コウが可愛い(((o(* *)o))) ”
2013 年 11 月 4 日,15:17.Tweet.
- ケイ (@kidousensi009)
“ SPEC 結～漸ノ編～おもろかたな 大島優子演技上手いな。 交ノ編早よ観たいな ”
2013 年 11 月 3 日,22:59.Tweet.

図 1 “大島優子”についてのダイナミックな意見が記述された tweet の例。

が更新される傾向がある。

ダイナミックな情報が含まれる tweet の例として、“大島優子”が含まれる tweet の一部を図 3.1 に示す。これらの tweet はユーザが“大島優子”が出演するテレビ番組や映画を視聴しながら感想や意見を投稿していると考えられる。また、tweet 中には“可愛い”“面白い”“演技上手い”などの評価情報が含まれていることがわかる。これらのことから、tweet 中の情報を用いることで「その瞬間」における対象知識についての感性評価情報を取得可能と考えられる。本稿では、対象知識に関する情報を tweet から取得することで、対象知識に関するダイナミックで感性的な情報の取得を図る。

4 提案手法

提案手法では、tweet から得られた情報を用いることで、対象知識と階層知識から得られた対象知識の階層的同位語との感性的な結びつきを可視化する。提案手法の基本的な処理手順を以下に示す。

Step1 対象知識および対象知識の階層的同位語をクエリとして Twitter 上のタイムラインで検索し、それぞれのクエリについて tweet を N 件を取得する。

Step2 取得した tweet を形態素解析し、評価、印象に影響すると思われる単語をそれぞれのクエリ毎に抽出する。

Step3 それぞれのクエリ毎に抽出した各単語の TF 値

を算出し、それぞれのクエリについての関連情報として扱う。

Step4 対象知識と対象知識の階層的同位語の関連情報について類似度を算出し、この類似度を感性的な結びつきとする。

4.1 関連情報の取得

対象知識および対象知識の階層的同位語を Twitter のタイムライン上で検索し、得られた tweet 中の単語頻度をそれぞれの知識に紐づく関連情報とする。本稿では、形態素解析によって得られた全ての単語から品詞が「名詞、一般」「名詞、固有名詞」「名詞、サ変接続」「形容詞、自立」「名詞、ナイ形容詞語幹」「動詞、自立」であるものを抽出する。これは、“の”“や”などの助詞や“しかし”“ゆえに”などの接続詞は客観的情報と評判情報のどちらとしても利用価値がなく、知識に紐づく関連情報としてはノイズとなってしまう可能性が高いためである。形態素解析器には、MeCab を用いた。ここで、形態素解析のための辞書には、Wikipedia の項目名を固有名詞として追加した。

形態素解析によって得られた抽出対象の品詞である単語について、それぞれ TF 値を算出する。そして、対象知識および対象知識の階層的同位語の TF 値それぞれについて Harman の正規化を適用する。正規化された TF 値 $tf(i, j)$ は下式で算出される。

$$tf(i, j) = \frac{\log_2(freq(i, j) + 1)}{\log_2(length_j)} \quad (1)$$

ここで、 $freq(i, j)$ はクエリ j についてのツイート中における単語 i の頻度を示し、 $length_j$ はクエリ j についてのツイート中における単語の種類数を示す。

4.2 対象知識と対象知識の階層的同位語との感性的結びつき算出

対象知識と対象知識の階層的同位語について式 1 で算出された正規化 TF 値を情報として、対象知識と対象知識の階層的同位語の類似度を算出する。このとき、対象知識および対象知識の階層的同位語は、各単語の $tf(i, j)$ を要素としてもつベクトルとして表現する。対象知識を \vec{Q} 、対象知識の階層的同位語 k を \vec{C}_k とすると、 \vec{Q} と \vec{C}_k の類似度は下式で算出される。

$$\cos(\vec{Q}, \vec{C}_k) = \frac{\sum_{i=1}^{|V|} Q_i C_{ki}}{\sqrt{\sum_{i=1}^{|V|} Q_i^2} * \sqrt{\sum_{i=1}^{|V|} C_{ki}^2}} \quad (2)$$

ここで、 $|V|$ は \vec{Q} および \vec{C}_k の tweet における単語種類数を示す。

5 評価実験

本節では、対象知識と対象知識の階層的同位語の感性的結びつきの算出実験を行う。実験結果の考察を通して、提案手法の有用性について議論する。

本稿では、隅田らによって提案された上位下位語抽出ツール [6] を利用することで対象知識の上位下位概念を獲得し、対象知識と共通の上位概念を持つ階層的同位語を得る。隅田らのツールを利用することにより、Wikipedia の記事構造に含まれる節や箇条書きの見出しから、多くの一般的知識について上位概念と下位概念が獲得可能となる。

本稿では、30 件の階層的同位語を評価実験の階層的同位語として用いる。TF 値の算出に用いる tweet の取得件数 N は 100 件とし、取得時のタイムラインから過去 24 時間以内に投稿された tweet のみを対象とする。tweet は 11 月 1 日と 11 月 6 日に取得している。対象知識は、“新垣結衣”“田中将大”“立命館大学”とし、それぞれの上位概念を“日本の女優”“スポーツ選手”“日本の大学”とした。

表 2, 3, 4 に、tweet 取得日毎の“新垣結衣”“田中将大”“立命館大学”とそれぞれの階層的同位語との感性的結びつきを示す。以下、5.1, 5.2 節でそれぞれ、感性的に適切な階層的同位語の抽出、感性的結びつきの変化について考察する。

5.1 対象知識の比較対象として適当な階層的同位語の抽出

表 2 の、対象知識である“新垣結衣”は、主にテレビドラマや映画、CM に出演し活躍している日本の女優である。同表から、“新垣結衣”と同様にテレビドラマや映画、CM に出演している日本の女優である“堀北真希”や“上戸彩”“佐々木希”などは感性的な結びつきが強いことが見て取れる。一方で、“小倉優子”や“スザンヌ”はドラマや映画ではなく主にバラエティ番組で活躍しており、一般的には女優としてよりもタレントとして知られている。これらの知識については、“新垣結衣”との感性的結びつきが弱い傾向にあることがわかる。本結果より、対象知識に対して直感的に類似した知識は類似度が高く、対象知識の比較対象として直感的に不適当な知識は類似度が低い傾向が見られた。tweet 中の単語を用いて対象知識と対象知識の階層的同位語との類似度を算出することで、感性的に比較可能な知識が獲得可能であることが示唆された。

表 3 で対象知識とした“田中将大”はプロ野球選手であり、投手として活躍している。同表より、プロ野球選手で投手として活躍している“松阪大輔”“上原浩治”“三浦大輔”との間には高い類似度が確認できる。また、“斎藤佑樹”は、実験に用いた階層的同位語の中

表 2 “新垣結衣”と階層的同位語の感性的結びつき。括弧内は結びつきの強度による順位。

比較可能な知識	コサイン類似度 (2013 年 11 月 1 日)	コサイン類似度 (2013 年 11 月 6 日)
堀北真希	0.482 (1)	0.643 (1)
上戸彩	0.461 (2)	0.396 (5)
佐々木希	0.456 (3)	0.375 (6)
宮崎あおい	0.448 (4)	0.396 (4)
戸田恵梨香	0.440 (5)	0.353 (9)
石原さとみ	0.436 (6)	0.403 (3)
北川景子	0.426 (7)	0.409 (2)
篠原涼子	0.416 (8)	0.351 (11)
岡本玲	0.410 (9)	0.302 (20)
吉高由里子	0.408 (10)	0.342 (13)
小池栄子	0.404 (11)	0.362 (7)
広末涼子	0.403 (12)	0.354 (8)
安田美沙子	0.402 (13)	0.278 (26)
押切もえ	0.396 (14)	0.275 (28)
水川あさみ	0.393 (15)	0.347 (12)
国仲涼子	0.390 (16)	0.279 (25)
観月ありさ	0.385 (17)	0.340 (14)
井上真央	0.384 (18)	0.264 (29)
山本梓	0.379 (19)	0.292 (23)
前田敦子	0.374 (20)	0.295 (21)
米倉涼子	0.369 (21)	0.321 (15)
大島優子	0.367 (22)	0.305 (18)
森カンナ	0.361 (23)	0.305 (17)
加藤あい	0.358 (24)	0.277 (27)
指原莉乃	0.357 (25)	0.282 (24)
天海祐希	0.356 (26)	0.309 (16)
スザンヌ	0.347 (27)	0.294 (22)
長澤まさみ	0.341 (28)	0.353 (10)
黒木メイサ	0.339 (29)	0.303 (19)
小倉優子	0.240 (30)	0.263 (30)

でも“田中将大”との感性的結びつきが強い。“田中将大”は“斎藤佑樹”と、高校生時代からの好敵手として有名である。そのため、多くの人が“田中将大”について思い浮かべるとき、同時に“斎藤佑樹”を連想することが多くなると考えられ、妥当な結果が得られたと考える。一方で、プロボクサーである“亀田興毅”“亀田大毅”や体操選手である“田中理恵”、フィギュアスケート選手である“織田信成”などはスポーツ選手ではあっても、プロ野球選手でない。これらの知識と“田中将大”の類似度は比較的低い傾向にあり、直感的にも妥当な結果が得られた。

一般的に、“立命館大学”は関西の私立大学である“同志社大学”“関西学院大学”“関西大学”と比較されることが多い。表 4 においても、これらの知識に対しては類似度が高く、直感的にも妥当な結果と考える。

5.2 感性的結びつきの時間的变化

表 2 では、“岡本玲”を除いて結びつきの強度について上位の知識の多くが、取得日が変化したとしても大きな変動は見られなかった。取得日が変化したとしても、“新垣結衣”と同様に女優として CM やテレビ、ドラマに出演する女優は“新垣結衣”と感性的に近い知識とし

表3 “田中将大”と階層的同位語の感性的結びつき．括弧内は結びつきの強度による順位．

階層的同位語	コサイン類似度 (2013年11月1日)	コサイン類似度 (2013年11月6日)
斎藤佑樹	0.426 (1)	0.379 (1)
松坂大輔	0.387 (2)	0.341 (4)
星野仙一	0.377 (3)	0.326 (7)
上原浩治	0.369 (4)	0.348 (3)
三浦大輔	0.358 (5)	0.304 (14)
前田健太	0.337 (6)	0.330 (6)
松井秀喜	0.337 (7)	0.355 (2)
坂本勇人	0.332 (8)	0.265 (22)
内村航平	0.325 (9)	0.306 (12)
前田智徳	0.320 (10)	0.293 (16)
イチロー	0.313 (11)	0.302 (15)
長嶋茂雄	0.312 (12)	0.256 (24)
鳥谷敬	0.309 (13)	0.282 (18)
内海哲也	0.305 (14)	0.197 (30)
吉田沙保里	0.301 (15)	0.258 (23)
三浦知良	0.300 (16)	0.279 (20)
桧山進次郎	0.299 (17)	0.281 (19)
北島康介	0.296 (18)	0.323 (8)
石川遼	0.291 (19)	0.290 (17)
香川真司	0.286 (20)	0.306 (13)
木村沙織	0.286 (21)	0.317 (9)
本田圭佑	0.285 (22)	0.337 (5)
錦織圭	0.284 (23)	0.271 (21)
高橋大輔	0.284 (24)	0.252 (26)
浅尾美和	0.274 (25)	0.312 (10)
立浪和義	0.265 (26)	0.254 (25)
亀田興毅	0.264 (27)	0.307 (11)
織田信成	0.247 (28)	0.210 (28)
田中理恵	0.242 (29)	0.235 (27)
亀田大毅	0.204 (30)	0.208 (29)

て抽出されている．一方で，取得日の違いによって“岡本玲”と“長澤まさみ”は大きく順位が変動した．11月1日には“新垣結衣”と“岡本玲”を比較した tweet が多く存在したが，11月6日にはそれらの tweet は見られなくなった．そのため，11月6日には“岡本玲”は“新垣結衣”との感性的結びつきが低くなっている．また，11月6日の“長澤まさみ”についての tweet では，長澤まさみが主演する映画に関する記述が多く見られた．同日には，“新垣結衣”についての tweet でも映画に関する意見や感想が投稿されており，“長澤まさみ”と“新垣結衣”は感性的に近い知識として順位が上昇したと考えられる．

表4に示した“立命館大学”と“立命館大学”の階層的同位語との感性的結びつきでは，11月1日には“びわこ成蹊スポーツ大学”と“立命館大学”の類似度が0であったのに対して，11月6日には大きく類似度が跳ね上がり，“立命館大学”に対する結びつきが強くなっていることが見て取れる．取得日2013年11月1日には“びわこ成蹊スポーツ大学”についての tweet がタイムライン上に存在しなかったのに対して，取得日2013年11月6日になると“びわこ成蹊スポーツ大学”についての tweet がタイムラインに出現した．これは，2013年

表4 “立命館大学”と階層的同位語の感性的結びつき．括弧内は結びつきの強度による順位．

階層的同位語	コサイン類似度 (2013年11月1日)	コサイン類似度 (2013年11月6日)
同志社大学	0.361 (1)	0.334 (3)
龍谷大学	0.348 (2)	0.318 (8)
南山大学	0.345 (3)	0.327 (5)
早稲田大学	0.342 (4)	0.318 (9)
関西大学	0.342 (5)	0.341 (1)
法政大学	0.322 (6)	0.316 (2)
神戸大学	0.322 (7)	0.314 (10)
関西学院大学	0.321 (8)	0.300 (14)
日本体育大学	0.317 (9)	0.264 (21)
京都大学	0.316 (10)	0.286 (18)
筑波大学	0.306 (11)	0.324 (6)
中央大学	0.305 (12)	0.332 (4)
慶應義塾大学	0.305 (13)	0.289 (17)
近畿大学	0.305 (14)	0.299 (17)
京都産業大学	0.304 (15)	0.301 (15)
東京理科大学	0.301 (16)	0.302 (12)
東海大学	0.299 (17)	0.285 (19)
お茶の水女子大学	0.298 (18)	0.292 (16)
一橋大学	0.296 (19)	0.283 (20)
大阪大学	0.313 (20)	0.264 (22)
岡山商科大学	0.287 (21)	0.225 (27)
北海道大学	0.286 (22)	0.322 (7)
上智大学	0.286 (23)	0.312 (11)
甲南大学	0.285 (24)	0.248 (24)
摂南大学	0.280 (25)	0.261 (23)
立命館アジア太平洋大学	0.266 (26)	0.188 (29)
びわこ学院大学	0.230 (27)	0.000 (30)
滋賀医科大学	0.224 (28)	0.239 (25)
いわき明星大学	0.095 (29)	0.216 (28)
びわこ成蹊スポーツ大学	0.000 (30)	0.234 (26)

11月9日，10日に“立命館大学”の近隣のショッピングセンターで“びわこ成蹊スポーツ大学”についての催しが開かれており，宣伝のための tweet 中に出現する単語が“立命館大学”と共有されたためと考える．本結果は，Tweet がもつダイナミック性の効果が確認できた良例であると言える．

6 おわりに

本稿では，tweet 中の情報を利用することで，対象知識と対象知識の階層的同位語に対して感性的な結びつきを算出する手法を提案した．提案手法では，Twitter がもつダイナミック性を利用することで，感性的結びつきの時間的に可視化する．

提案手法を用いて評価実験を行ったところ，対象知識に対して印象が類似した階層的同位語は感性的結びつきが強くなる一方で，直感的に比較対象として不適当と考えられる知識は結びつきが弱くなる傾向が見られた．また，感性的結びつきの変化に寄与した情報を考察することで，対象知識と対象知識の階層的同位語の感性的結びつきの様子が変化した要因について議論した．その結果，

tweet 取得日によって変化するタイムラインの情報を用いることで、時間的に変化する知識の関連情報が更新され、知識間の感性的な結びつきが時間経過に伴って変化することが確認された。これらのことから、提案手法が階層的な同位語の中でも直感的により適当な比較対象を抽出すると同時に、時間的に変化する知識間の関係性を可視化可能であることが示唆された。

本稿で知識についての関連情報を Twitter は 140 字の文字制限があるため、正式名称ではなく略称や愛称によって記述されることが多い。山西らの研究では、Web を情報源とする場合、正式名称だけでなく、愛称や略称などの別称を用いることで、より感性的な意見が得られる可能性が高いことを示している [7]。今後は、正式名称が含まれる tweet 中の情報だけでなく、tweet 中の愛称や略称などの情報にも着目することにより、今以上に感性情報を獲得し、感性的に比較可能な知識の獲得手法の構築を目指す。

謝辞

本研究は一部、中部電気利用基礎研究振興財団の助成のもと行われた。

参考文献

- [1] 大島裕明, 小山聡, 田中克己: “Web 検索エンジンのインデックスを用いた同位語とそのコンテキストの発見”, 情報処理学会論文誌, vol. 47(19), pp. 98–112 (2006).
- [2] 服部峻, 大島裕明, 小山聡, 田中, 克己: “継承関係と同位関係に基づく概念階層の Web からの抽出”, 日本データベース学会 Letters, vol. 6(2), pp. 9–12 (2007).
- [3] R. Yamanishi, J. Fukumoto, and F. Masui: “Semantical-coordinate Term Detection from Hierarchical Knowledge Using Web Snippets”, *Procedia Computer Science*, vol. 22, pp. 1276–1284 (2013).
- [4] J. Bollen, H. Mao and X. Zeng: “Twitter mood predicts the stock market”, *Journal of Computational Science*, vol. 2(1), pp. 1–8 (2011).
- [5] D. Ramage, S. Dumais, and D. Liebling: “Characterizing Microblogs with Topic Models”, in *Proc. of the Fourth International AAAI conference on Weblogs and Social Media*, pp. 130–137 (2010).
- [6] 隅田飛鳥, 吉永直樹, 鳥澤健太郎: “Wikipedia の記事構造からの上位下位関係抽出”, *自然言語処理*, vol. 16(3), pp. 3–24 (2009).
- [7] 山西良典, 福本淳一: “Wikioedia の特徴表現を利用した別称コーパス生成ツールの開発”, ARG 第 3 回ウェブインテリジェンスとインタラクション研究会 (2013) (to appear).