

# 機械学習による単語概念の意味属性推定

長谷川 美夏      小林 哲則      林 良彦

早稲田大学理工学術院

mika@pcl.cs.waseda.ac.jp

## 1 はじめに

単語が指示する概念(単語概念)を表現する手段として、その性質や特徴を表す属性<sup>1</sup>の集合による手段が componential approach として広く知られている [1]. 属性を獲得・構造化する方法として、人による評定を収集する方法がある. このようにして収集されたデータは、意味属性規範 (semantic feature norms; 以下、意味属性) などと呼ばれ、人間が保持している知覚・経験に基づく意味知識を明示化したものとして様々な研究において利用されている [2-5]. しかし、人間による評定を収集して構成される意味属性データは収集コストが高いため、対象となる単語概念のカバー率が低く、単語概念に付与される意味属性の完備性も十分でないといった問題がある. これらの問題点の解決は意味属性の適用範囲の拡大につながり、例えば単語概念間の共通性・差異性を捉えるために有用である.

本論文では、McRae らにより作成された意味属性データセット (以下、McRae データセット) [6] から単語概念と意味属性の対応関係を機械学習を用いて学習する. また、これに基づき未知の単語概念に対する意味属性を推定する手法を提案し、同データセットを用いた評価実験の結果を示す.

## 2 McRae データセットと関連研究

### 2.1 McRae データセット

McRae データセットは、基本レベル (basic level) に属すると考えられる 541 種類の英語単語概念に対し、700 名以上という多数の評定者が与えた総計 7,526 件の意味属性データを収集したものである. 2,526 種類の意味属性が用いられているが、本研究ではこれらを相互に独立なものとして扱い、単語概念をベクトル

<sup>1</sup>英語では、attribute, feature, property, component など様々な用語が使われているのが現状である.

表現する際の次元として扱う. 本データセットには各単語概念の規定・記述に関する豊富な情報が収録されているが、本研究では各単語概念における各意味属性に重みを与えるために、何人の評定者がそれぞれの意味属性データを付与したかの頻度を利用する.

### 2.2 関連研究

テキストコーパスから抽出する単語の分散表現の有用性は広く認識され、様々な意味タスクにおける有用性が確立している. しかしながら、人間が非明示的にもっている知覚や経験に基づく意味知識を取り入れた意味表現を求める動きも盛んである. とくに、画像から得られる視覚特徴を統合的に利用することで、意味的類似度・関連度のタスクで優れた推定精度が得られることが報告されている [7-10].

McRae データセットは、このような知覚・経験情報を人間による評定によって明示的に抽出したものであるが、概念のカバー率や意味属性の完備性の問題がある. データセットでカバーされていない単語概念に対しても意味属性を推定することにより、概念のカバー率を向上させようとする代表的な研究例として、[4,5] がある. とくに [4] は、意味属性と視覚空間の間で学習したクロスモーダルマップを用いて、テキストではなく視覚特徴から意味属性を推定する方法を提案しており、視覚特徴の有効性を示している. また [5] は、McRae データセットが人間による評定実験によって作成されたことに起因する意味属性の不完備性を指摘し、この解決の重要性を主張している.

## 3 機械学習による意味属性推定

前述のように、本研究においては単語概念の意味は意味属性の集合によって表される. よって、単語概念の意味属性の推定は、ある形式で表現された単語を、その単語を表現する意味属性の集合へとマッピングす

る過程とみなすことができる。本研究では、各種の手法により求められたベクトル  $x$  により単語を表現し、McRae データセットにおける意味属性に対する重みを各次元として持つベクトル  $y$  により意味属性の集合を表現する。また、マッピング  $y = f(x)$  の手法として、統計的回帰手法とニューラルネットを比較する。

### 3.1 単語概念の表現

単語が指示する概念の意味特徴は、一義的には言語産出の結果であるコーパスから抽出することが可能である。また、その知覚的な側面に関する特徴は、単語概念と対応付けられた画像データから抽出することが期待できる。そこで、以下の特徴量を比較する。

- **Ling\_EMBED**: 単語概念がもつ一般的な意味特徴が反映されていることを期待する特徴量。Word2Vec の CBOW モデルにより Google News コーパス (3 億単語) から学習された語彙数 300 万語の 300 次元の分散表現ベクトル<sup>2</sup>。
- **Ling\_TFIDF**: とくに単語概念の意味属性に関する特徴が反映されていることを期待する特徴量。ある単語に対する WordNet における定義文 (複数語義をもつ場合はこれらを連結) と、この単語を見出しに持つ Wikipedia 記事の要約文を連結したものを文書とみなし、WordNet に収録されている名詞 65,488 語に対して構成した文書集合から、各文書における要素単語に対して産出した TF/IDF 値を要素とするベクトルにより、各単語に表現を与えた。以下に述べる実験では、10,027 単語を要素として用いたので、ベクトルの次元数は 10,027 である。
- **Vis\_CNN**: 単語概念がもつ一般的な知覚的 (視覚的) 特徴が反映されていることを期待する特徴量。各単語をクエリとする Bing 画像検索による上位 10 件の画像について、ImageNet で学習済みの GoogLeNet モデル<sup>3</sup> を用いて特徴量を抽出し、次元ごとに平均を取った 1,024 次元のベクトル。
- **Vis\_CNN+RGB**: 色に関する意味属性が多く現れることを想定し、Vis\_CNN の特徴量に 768 次元の RGB ヒストグラムを連結した 1,792 次元のベクトル。

<sup>2</sup><https://github.com/mmhaltz/word2vec-GoogleNews-vectors>

<sup>3</sup>[https://github.com/BVLC/caffe/tree/master/models/bvlc\\_googlenet](https://github.com/BVLC/caffe/tree/master/models/bvlc_googlenet)

### 3.2 意味属性集合の表現

意味属性ベクトルの各次元は対応する意味属性の重みを表す。以下の 2 つの手法による重みを比較する。

- **features**: McRae データセットにおいて、30 人の被験者のうち何人がその feature をラベル付けたかの頻度により重みを与える。
- **weighted\_features**: 対象の単語概念について他の単語概念と比較したとき特徴的となる意味属性に大きな重みを与えるために、単語概念における意味属性の頻度をその意味属性の全単語概念における割合に乗じた結果を用いる。例えば、'apple' における意味属性 'is\_red' の頻度は 26、この意味属性の頻度合計は 382 であるため、weighted\_features における重みは、 $26 \times \frac{26}{382} = 1.76$  となる。

### 3.3 マッピング手法

以下の回帰手法、ニューラルネットワークを用いる手法を比較する。

- **PLSR**: 先行研究 [4] との比較を行うため、部分最小二乗回帰 (partial least square regression) 法を用いる。ハイパーパラメータである潜在変数の次元数は 50 とした。
- **ReLU**: 1 層のネットワークで線形変換  $y = W \cdot x + b$  を行う。式 1 の Leaky ReLU ( $a=0.2$ ) を活性化関数として用いる。損失関数は式 2 の最小二乗誤差を用いる。

$$f = \begin{cases} x & (x > 0) \\ 0.2x & (x \leq 0) \end{cases} \quad (1)$$

$$Loss = \min \sum_{t=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

- **MLP (Multi Layer Perceptron)**: 3 層の MLP によりネットワークを構成する。活性化関数、損失関数は上記 ReLU と同様である。

## 4 実験

### 4.1 実験設定

図 1 に示す過程で、推定された意味属性ベクトルが当該の単語の意味表現として適切であるかを評価する。このために、単語概念の表現 (4 種類を比較) を入力とし、2,526 次元の意味属性ベクトルを出力として求め、McRae データセットから求めた意味属性ベク

表 1: 各入力特徴量からの推定した意味属性ベクトルの評価

		features				weighted_features			
		P@1	P@5	P@10	P@20	P@1	P@5	P@10	P@20
RANDOM		0.17	0.97	1.88	3.70	0.17	0.97	1.88	3.70
PLSR	Ling_EMBED	2.76	18.66	36.27	54.41	0.0	1.2	4.33	9.41
	Ling_TFIDF	0.36	6.74	13.4	25.36	0.0	1.64	4.53	8.65
	Vis_CNN	0.69	9.54	18.33	30.19	0.0	0.58	1.06	2.7
	Vis_CNN+RGB	1.39	7.66	16.39	28.38	0.0	0.22	0.88	2.1
	Fagarasan et al. [4]	<b>2.90</b>	13.04	24.64	49.28	-	-	-	-
ReLU	Ling_EMBED	2.39	23.44	39.87	58.32	0.0	4.84	10.94	17.72
	Ling_TFIDF	1.98	15.14	26.05	41.08	<b>0.58</b>	<b>8.31</b>	<b>14.88</b>	22.04
	Vis_CNN	0.81	11.48	20.3	31.63	0.0	0.44	1.48	3.33
	Vis_CNN+RGB	0.22	4.37	10.0	18.44	0.0	1.33	2.81	6.89
MLP	Ling_EMBED	2.83	26.34	<b>45.39</b>	<b>62.35</b>	0.0	2.97	6.4	15.1
	Ling_TFIDF	1.38	<b>27.97</b>	43.05	58.68	0.0	6.73	11.52	<b>22.91</b>
	Vis_CNN	0.52	10.81	19.19	29.78	0.0	0.81	1.48	3.19
	Vis_CNN+RGB	1.19	9.93	19.11	31.04	0.0	0.44	1.19	3.63

トル (以下, 正解意味属性ベクトル) とのコサイン類似度により近傍単語のランキングを導出する (図中の cosine similarity の箱). このとき, 上位  $k$  件における精度  $P@k$  を評価指標として用いる. 実験では, 541 概念について 4 分割交差検定 (学習 406 概念, テスト 135 概念)<sup>3</sup>を行った.

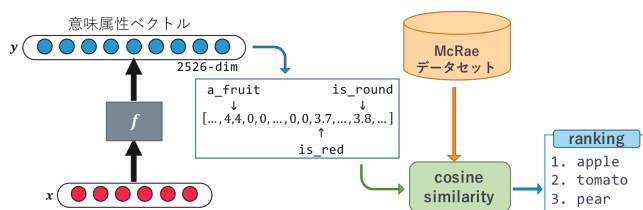


図 1: 意味属性推定プロセス図

## 4.2 実験結果

表 1 に各入力特徴量を用いた場合の  $P@k$  の結果を示す. 提案手法は  $P@1$  (features) 以外の結果において比較対象の先行研究 [4] を上回る精度を示した. なお [4] では, Wikipedia コーパスのうち頻度が上位の 1 万単語についての共起行列を PPMI と SVD で 300 次元に圧縮した言語ベクトルを PLSR を用いて features による意味属性ベクトルを求めている. features による意味属性ベクトルは, Ling\_EMBED に MLP を用いたときが最も推定精度が良く, weighted\_features によるベクトルは Ling\_TFIDF に ReLU を用いたときに最も良い推定精度が得られた.

## 4.3 共通性と差異性を捉える意味属性

表 2 に示すように, features による意味属性ベクトルは, 上位概念に関する意味属性 (a\_fruit, a\_bird, な

<sup>3</sup> [4] の実験設定に合わせた. [5] では, one leave out による交差検定が行われている.

ど) が推定されやすい傾向にある. 従って, 近傍単語の探索の精度が良い.

一方, weighted\_features による意味属性ベクトルは他の概念との差異を捉える固有な意味属性の推定を目的としている. 例えば表 2 を見ると ReLU(Ling\_TFIDF)  $\rightarrow$  weighted\_features の結果では 'apple', 'banana' は共に推定結果に 'is\_purple' が現れてしまっている. これは, 学習データに含まれていた 'grape', 'plum' の影響を強く受けているといえる.

表 2 において\*が付いている意味属性は, McRae データセットには含まれていない結果であるが, これらには必ずしも不適切とはいえないものが含まれる. 例えば, 'dove' の推定結果として lives\_in\_cities\*, lives\_in\_parks\*, beh\_.\_coos\* などがある. これらは 'dove' を特徴付ける属性とみなすことができる. 以上から, weighted\_features による意味属性ベクトルは概念を特徴付ける意味属性を抽出する上では有用である可能性がある.

## 4.4 意味属性の補完可能性

以上の結果から示唆されることは, McRae データセットにおける意味属性の不完備性 [5] であるが, 人手による評定によるリソースである限りこの問題は避けられない. しかし, 提案手法のような機械的な手段によって自動的な補完できれば, 意味属性データの利用範囲の拡大につながる.

そこで, 推定された意味属性のなかにどの程度, 適切と考えられる意味属性が含まれるかの調査を本論文の筆頭著者が行った. ランダムにサンプルした 100 単語概念に対して, それぞれ上位 10 件にランクされた意味属性の総計 1,000 件から McRae データセットに含まれる 155 件を除いた 845 件に対して, 意味属性が適切:y, 場合によっては適切:s, 不適切:n の 3 段階

表 2: 推定された意味属性 (上位 5 件)

	MLP(Ling_EMBED) → features	ReLU(Ling_TFIDF) → weighted features
apple	a_fruit, is_red tastes_sweet, is_round, has_seeds	is_purple*, has_a_pit* eaten_in_pies, a_fruit, has_seeds
banana	a_fruit, is_green tastes_sweet*, is_small*, is_edible	is_purple*, a_fruit has_seeds*, has_a_pit*, tastes_sweet*
dove	a_bird, is_green* beh_-_flies, has_feathers, found_in_bathrooms*	lives_in_cities*, lives_in_parks* beh_-_coos*, beh_-_produces_manure*, has_2_feet*
sparrow	a_bird, beh_-_flies has_feathers, has_wings, has_a_beak	has_a_long_beak*, beh_-_puts_holes_in_trees* has_a_strong_beak*, a_bird, found_in_cartoons*
skirt	worn_by_women, is_long clothing, made_of_material, made_of_silk*	worn_for_weddings*, is_feminine* is_formal*, worn_by_girls*, different_styles*
shirt	clothing, different_colours worn_for_warmth, is_long*, made_of_wool*	eg_-_a_pull-over*, has_sleeves made_of_wool*, worn_for_winter*, a_shirt*

表 3: 最近傍単語 (上位 5 件)

	MLP(Ling_EMBED) → features	ReLU(Ling_TFIDF) → weighted features
apple	cherry, <b>apple</b> , tomato, strawberry, plum	cherry, eggplant, plum, strawberry, tangerine
banana	pear, apple, plum, pineapple, avocado	pineapple, tangerine, eggplant, cherry, nectarine
dove	partridge, sparrow, pheasant, nightingale, blackbird	pigeon, blackbird, nightingale, pheasant, raven
sparrow	starling, finch, <b>sparrow</b> , blackbird, chickadee	woodpecker, pigeon, blackbird, bat_(animal), nightingale
skirt	gown, dress, <b>skirt</b> , blouse, camisole	dress, veil, gown, sweater, blouse
shirt	sweater, coat, shawl, <b>shirt</b> , jacket	sweater, blouse, coat, scarf, swimsuit

で評定を行った。その結果、y のみを正解と考える場合 178 件 (21.1%)、y+s を正解と考える場合 259 件 (30.7%) が適切であるという結果を得た。現状での適切な意味属性の含有率は低いが、今後、フィルタリング・選出手法を検討する価値はあるものとする。

## 5 おわりに

人手により評定された意味属性には人間の持つ非明示的な意味知識が含まれ利用価値が高いが、これを自動的に獲得することは挑戦的な課題である。本論文では、ベクトルで表される単語特徴量を入力とし、意味属性のベクトル表現を出力とする機械学習手段により単語概念の意味属性を推定する手法を提案し、評価実験により、提案手法が既存研究と同等以上の精度を持つことを示した。また、データセットには含まれていないが妥当と考えられる単語概念の意味属性を一定の精度で推定できることを示した。実際に有効な意味属性を選出するためには、単語概念の共起の偏りを利用することが有効であると考え、現在、検討を進めている。未知の概念について概念間の差異を特徴付ける意味属性を機械学習を用いて推定するのは困難であり、概念間の差異を得るには既存の知識リソースの活用が必要であると考えられる。

## 謝辞

本研究は JSPS 科研費 (17H01831) の助成を受けた。

## 参考文献

- [1] M.L. Murphy. *Lexical Meaning*. Cambridge Textbooks in Linguistics. Cambridge University Press, 2010.
- [2] Carina Silberer, Vittorio Ferrari, and Mirella Lapata. Models of Semantic Representation with Visual Attributes. *Assoc. Comput. Linguist.*, pp. 572–582, 2013.
- [3] 林良彦. Web 画像による語義・概念の視覚的な提示に関する検討. 人工知能学会第 5 回インタラクティブ情報アクセスと可視化マイニング研究会, pp. 19–26, 2013.
- [4] Luana Fagarasan, Eva Maria Vecchi, and Stephen Clark. From distributional semantics to feature norms : grounding semantic models in human perceptual data. *Proc. 11th Int. Conf. Comput. Semant. London, UK, April 15–17 2015*, pp. 52–57, 2015.
- [5] Luana Bulat, Douwe Kiela, and Stephen Clark. Vision and Feature Norms : Improving automatic feature norm learning through cross-modal maps. *Proc. NAACL-HLT 2016*, pp. 579–588, 2016.
- [6] Ken McRae, George S Cree, Mark S Seidenberg, and Chris McNorgan. Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods*, Vol. 37, No. 4, pp. 547–559, 2005.
- [7] Elia Bruni, Daniel Gatica-perez, Nam Khanh Tran, and Marco Baroni. Multimodal distributional semantics. *J. Artif. Intell. Res.*, Vol. 49, No. December, pp. 1–47, 2014.
- [8] Carina Silberer and Mirella Lapata. Learning Grounded Meaning Representations with Autoencoders. *Proc. 52nd Annu. Meet. Assoc. Comput. Linguist. (Volume 1 Long Pap.)*, pp. 721–732, 2014.
- [9] Angeliki Lazaridou, The Pham Nghia, and Marco Baroni. Combining Language and Vision with a Multimodal Skip-gram Model. *NAACL 2015*, pp. 153–163, 2015.
- [10] Mika Hasegawa, Tetsunori Kobayashi, and Yoshihiko Hayashi. Incorporating visual features into word embeddings : A bimodal autoencoder-based approach. *IWCS2017*, 2017.