

概念間の想起の強さと方向性の予測

Predicting the Strength and Directionality of Evocation Relations

林 良彦

Yoshihiko Hayashi

早稲田大学理工学術院

Faculty of Science and Engineering, Waseda University

Evocation is a directed yet weighted semantic relation between concepts. This paper presents a supervised-learning approach to predict the strength and to determine the directionality of a potential evocation relation. Empirical results that investigated useful features along with effective machine learning frameworks are shown, indicating that the combination of the proposed features largely outperformed individual baselines, and also suggesting that *semantic relational vectors* computed from existing semantic vectors for lexicalized concepts were indeed effective in both of the prediction of strength and the determination of directionality.

1. はじめに

想起関係 (evocation) とは, ある概念 (起点概念 s) がどの程度ほかの概念 (ターゲット概念 t) を心理的に思い起こさせる (bring to mind) か [2] を表す概念間の意味的な関係であり,

- 同一の品詞の間には限らない,
- 方向性のある非対称な関係であり,
- 想起の度合い (強さ) を表す重み付きの関係である,

という特徴を持つ.

想起関係は, 人間の知覚や経験に基づく意味関係 [8] であることから, 言語産出の結果を集積したコーパスからは得ることができない, あるいは, 得ることが難しい. また, 想起の機序やそのタイプも必ずしも明らかではないため, 任意の概念ペアの間の想起関係の有無, 方向性, その強さを予測することは一般的には困難な問題である. しかし, 与えられた概念ペアに対して想起関係を精度よく予測する機械的な手段が実現できれば, 様々な意味的な言語処理タスクに適用可能であると考えられる.

本論文は, 想起関係の強さの予測と方向性の判定を教師付き機械学習により行う方法を提案する. より具体的には, 想起関係の強さの予測は回帰問題, 想起の方向性の判定は分類問題として扱い, 様々な属性の有効性を実験的に検討する.

なお, 本研究で扱う想起関係のデータは, 後述するように Princeton WordNet (PWN) [10] における語彙化概念 (lexicalized concept) の間の想起関係を取り扱っている. すなわち, 本研究で扱う想起関係は英語の語彙化概念の間の関係である.

2. 想起関係データ

想起関係の強さと方向性の双方が系統的に付与されたデータは存在しない. このため, それぞれのアノテーションが付与された以下の2つのデータセットを用いる.

2.1 Princeton WordNet の想起データ

想起関係の強さが付与されたデータとして, Princeton WordNet の研究グループにより構築・公開されているデータ^{*1}[2]を用いる.

連絡先: 林 良彦, 早稲田大学理工学術院・実体情報学博士プログラム. mailto:yshk.hayashi@gmail.com

*1 <http://wordnet.cs.princeton.edu/downloads.html>

このデータは, PWN においてコアとなる 1,000 の語彙化概念 (core synsets) からランダムに抽出された 119,652 件の語彙化概念ペアの間の想起の程度を, 少なくとも 3 名の被験者に評定させたものである. 想起の強さは 0 から 100 の間の評定値により表現されるが, 何らかの心的な関連性があることを意味する 0 より大きい評定値が付与されたペアは, 39,309 件 (全体の約 33%) にとどまっている. また, x から y , y から x というように双方向の評定値が付与されているペアの数は 7,164 ペアであり, これには少なくともどちらかの評定値がゼロであるものも含む.

[2] では, 想起の強さと各種の尺度による意味的類似度・関連度の間の相関が低いことが報告されており, 想起関係が通常の意味的類似性・関連性とは異なるタイプの情報であり, その予測は困難を伴うことが議論されている. また, LSA を用いた関連度と想起強度との Spearman 相関係数は 0.131 であったと報告されている.

2.2 Ma による evocationNet データ

想起関係の方向性が付与されたデータとして, Ma [8] による evocationNet データ^{*2}を用いる.

このデータは, 単語間の自由連想に関するデータ^{*3}をベースとし, 対象単語の語義解消を行うことにより, 語彙化概念間の想起関係へと変換したデータである. 公開されている evocationNet データでは, 348,447 (synset レベルでの重複含む) の synset ペア (x, y) に対し, 表 1 に示すような方向性に関する分類が付与されている.

表 1: Ma の evocationNet データにおける方向性種別の分布.

Directionality	Count
$x \rightarrow y$ (outbound)	172,126
$x \leftarrow y$ (inbound)	123,147
$x \leftrightarrow y$ (bidirectional)	43,459
no-evocation	9,715
Total	348,447

*2 <http://kettle.ubiq.cs.cmu.edu/~xm/DataSet/webpage/evocationNet/>

*3 University of South Florida Free Association Norm Data. <http://w2.usf.edu/FreeAssociation/>

3. 機械学習による想起関係の予測

想起関係の背後にある心理的な機序は明確でないが、様々な要因が相互に関係していると考えられる [5] ため、寄与の可能性が考えられる様々な属性を考慮した機械学習により、その強さの予測 (回帰問題)、方向性の決定 (分類問題) を行うことが妥当であると考えられる。

本報告では、双方の問題において用いる属性集合は共通とし、評価実験において問題に応じた属性の有効性を実験的に検討する。また、機械学習の枠組みとして、ランダムフォレスト (RF) とフィードフォワード・ニューラルネットワーク (NN) の2つを比較する。以下では、機械学習において利用する属性を3つの属性群に分けて説明する。

3.1 類似度・関連度に基づく属性

想起関係は非対称な意味関係であるが、そのある部分は、対称性のある意味的な類似度・関連度により基盤付けられている、あるいは、バイアスされていると考えられる [5]。そこで、以下に示す4つの意味的類似度・関連度を属性とする。これらの属性はいずれも一次元のスカラー値である。

この中の2つ ($ldaSim$, $w2vSim$) は単語間の意味的類似度・関連度である。 $ldaSim$ は従来からの分布意味論に基づく手法として、 $w2vSim$ は近年注目を浴びている分散表現に基づく属性として採用した。本研究の対象は概念間の関係であるが、PWN 想起データの場合は、synset を規定する代表単語、Ma の evocationNet データの場合は、データ作成の基となった単語を利用することで、単語間の意味的類似度・関連度を考慮することができる。

他の中のもの2つ ($wupSim$, $autoexSim$) は概念間の意味的類似度・関連度である。 $wupSim$ は既存の語彙資源の構造に基づく属性、 $autoexSim$ は分散表現に基づく属性として採用した。

- $ldaSim$ は、LDA 手法により導出された単語ベクトル間のコサイン類似度を与える。今回は、enwik9 と呼ばれるコーパス*4 から単語ベクトルを導出した。
- $w2vSim$ は、Word2Vec 手法 [9] により導出された単語ベクトル (W2V ベクトル) の間のコサイン類似度を与える。今回は、Google's Word2Vec site*5 で公開されている単語ベクトルを利用した。これは、CBOW (Continuous Bag-Of-Words) モデルにより導出された 300 次元のベクトルである。
- $wupSim$ は、語彙資源 PWN に基づく概念間 (synset 間) の意味的関連度である Wu-Palmer による関連度 [3] を与える。この指標は、各品詞に対応するネットワークを統合するルートノードを仮定することにより、品詞の異なる概念間の関連度を求めることが可能としており、今回の想起データに適用するのに好都合である。
- $autoexSim$ は、AutoExtend と呼ばれる手法 [11] により導出された語彙化概念に対する分散表現ベクトル (AE ベクトル) の間のコサイン類似度を与える。今回は、提案者である Rothe のサイトで公開されているデータ*6 を用いた。この AE ベクトルは、上記の Word2Vec データから導出されており、各次元が W2V ベクトルと同じ意味合いを持つ 300 次元のベクトルである。

注 1) AutoExtend 手法 [11]: 単語の分散表現ベクトルを入力とし、PWN などの語彙資源が与えるネットワーク構造 (単語・語義・語彙化概念をノードとする) に対応する自己符号化器を適用することにより、語義 (正確には lexeme)、語彙化概念 (synset) に対する分散表現ベクトルを導出する。

3.2 語彙資源に基づく属性

本研究では、PWN における語彙化概念間の想起関係を対象としているが、PWN のような語彙資源自体も属性の情報源として有用である。本研究では、以下の3つの PWN 由来の属性を用いている。

これらの属性はいずれも、想起関係における非対称性的一端を捉えることを目的として導入している。

- $posSem$ は、起点概念、ターゲット概念、それぞれの語彙化概念の基本的な性質を表す。より具体的には、各語彙化概念に対する品詞、粗いレベルの意味分類を表す 50 次元 (品詞:5, 意味分類:45) のバイナリベクトルをこの順に連結した 100 次元のベクトルである。ここで、粗い意味分類の体系としては、PWN における lexicographer file を用いている。
- $lexNW$ は、起点概念、ターゲット概念、それぞれに対応するノードが PWN が形成するグラフ構造においてどの程度重要であるかを表す指標を組にしたベクトルである。重要度の指標としては、betweenness centrality と load centrality [1] を用いている。
- $dirRel$ は、PWN のグラフ構造を利用して、非対称な意味的類似度における基本的な考え方である「属性包含」 (feature inclusion) を模倣する指標であり、以下の式により定義される。ここで、 $nb(x, k)$ は、概念ノード x から k ホップ以内で到達可能な概念ノードの集合である。今回報告する実験結果では、前実験の結果に基づき、 $k = 3$ とした。

$$dirRel(s, t, k) = \frac{|nb(s, k) \cap nb(t, k)|}{|nb(s, k)|}$$

注 2) 属性包含 [6]: 多くの属性を持つオブジェクト x と、より少ない属性を持つオブジェクト y があるとき、 $asim(x, y) < asim(y, x)$ という傾向があるとする考え方。ここで、 $asim(x, y)$ は、 x から y を見たときの類似度を表す。

3.3 概念間の意味関係ベクトル

単語の分散表現を用いる利点の一つとして、意味関係を含む単語間の関係が、分散表現ベクトルの差分に現れるということがある。そこで、各想起関係の特徴が反映されていることを期待し、次式で与えられる起点概念とターゲット概念に対する AE ベクトルの差分ベクトル (300 次元) $relVec(s, t)$ を属性として用いた。ここで、 $semVec(s)$ は、語彙化概念 s に対する AE ベクトルを表す。

$$relVec(s, t) = semVec(t) - semVec(s)$$

本報告で述べる実験においては、300 次元の概念間の意味関係ベクトルは、他の属性を並べたベクトルと単純に連結して用いている*7。

*4 <http://mattmahoney.net/dc/text.html>

*5 <https://code.google.com/p/word2vec/>

*6 <http://www.cis.lmu.de/~sascha/AutoExtend/embeddings.zip>

*7 ベクトルの次元圧縮は別途試行したが、必ずしも良好な結果は得られていないので、本報告では扱わない。

4. 評価実験

機械学習の手法として、Random Forest (RF), および、3層のフィードフォワード・ニューラルネットワーク (NN) (dropoutあり、活性化関数:ReLU) の2手法を回帰問題、分類問題の双方に適用した。なお、RFはscikit-learn^{*8}, NNはChainer^{*9}を用いて実装した。

4.1 回帰問題としての想起の強さの予測

PWN 想起データの全データ (119,652 件) を対象とし、5分割交差検定により想起強度の予測に関する評価を行った。評価指標は、回帰問題の評価の標準的な手順に従い、gold-dataと予測値との相関係数 (Pearson: r , Spearman: ρ) を用いる。なお、gold-data g についてはその値域が広い ([0,100]) ため、 $g' = \log(g + 1)$ なる対数変換を適用した。

4.2 分類問題としての想起の方向性の決定

Ma の evocationNet の全データ (348,447 件) を対象とし、5分割交差検定により想起強度の予測に関する評価を行った。評価指標は、分類問題における標準的な指標である、Precision/Recall/F1/Accuracy を用いる。

5. 実験結果と考察

5.1 想起の強さの予測

先に述べた全ての属性を用いることにより、NN 手法によって、 $r = 0.4391, \rho = 0.4000$ という結果を得た。この結果は、RF 手法による $r = 0.3695, \rho = 0.3291$ という結果をかなり上回った。そこで、以下の議論は NN 手法によるものに限る。

表 2 に、想起の強さに関する回帰問題の予測結果を示す。表において、 $+foo$ という表記は属性 foo のみを用いた結果 (baseline) を示すのに対し、 $-foo$ という表記は、全属性の中で属性 foo のみを除いた結果 (ablation test) を示す。

表 2: 想起の強さの予測結果.

Feature	r	ρ
<i>All</i>	0.4391	0.4000
<i>+ldaSim</i>	0.1536	0.1111
<i>-ldaSim</i>	0.4378	0.3994
<i>+w2vSim</i>	0.2472	0.1841
<i>-w2vSim</i>	0.4370	0.3991
<i>+wupSim</i>	0.0907	0.0663
<i>-wupSim</i>	0.4387	0.3997
<i>+autoexSim</i>	0.2395	0.1924
<i>-autoexSim</i>	0.4333	0.3962
<i>+posSem</i>	0.2442	0.2489
<i>-posSem</i>	0.4269	0.3837
<i>+lexNW</i>	0.1379	0.1211
<i>-lexNW</i>	0.4379	0.3999
<i>+dirRel</i>	0.0839	0.0622
<i>-dirRel</i>	0.4385	0.4000
<i>+relVec</i>	0.2931	0.2763
<i>-relVec</i>	0.3959	0.3534

この表に示す結果から以下のようなことが確認できる。

- 全ての属性を組み合わせる場合 (*All*) の結果が最も良く、単独でこれを超える結果をもたらす属性は存在しない。よって、属性は相補的に機能している。
- 分散表現に基づく意味的関連度 (*w2vSim*, *synSim*) はその比較対象 (*ldaSim*, *wupSim*) に対して顕著に良好な結果を示しており、想起関係のような漠然性や多様性が高い意味関係においても、そのある部分を確実に捉えることができる。
- 概念間の意味関係ベクトル (*relVec*) を利用しないと、著しく結果が悪くなる ($r = 0.3959, \rho = 0.3534$) ことから、AE ベクトルに基づく意味関係ベクトルの有効性は高い。すなわち、分散表現ベクトルの差分により意味関係の特徴 (のある部分) を捉えられるという期待が裏付けられた。
- 対象とする概念ペアの基本的な性質を表す *posSem* の寄与度も当初の予想以上に高いことから、概念の基本的な性質の順序付き組合せも有用な手がかりとなる。

5.2 想起の方向性の決定

先に述べた全ての属性を用いることにより、RF 手法によって、*Accuracy* = 0.8548 という結果を得た。この結果は、NN 手法による *Accuracy* = 0.7642 という結果をかなり上回る^{*10}。そこで、以下の議論は RF 手法によるものに限る。

表 3: 想起の方向性の決定の結果 (*Accuracy*)

Feature	<i>Accuracy</i>
<i>All</i>	0.8548
<i>+ldaSim</i>	0.5065
<i>-ldaSim</i>	*0.8552
<i>+w2vSim</i>	0.4816
<i>-w2vSim</i>	0.8538
<i>+wupSim</i>	0.4938
<i>-wupSim</i>	*0.8550
<i>+autoexSim</i>	0.3496
<i>-autoexSim</i>	0.8544
<i>+posSem</i>	0.5290
<i>-posSem</i>	*0.8553
<i>+lexNW</i>	0.6885
<i>-lexNW</i>	0.8467
<i>+dirRel</i>	0.5229
<i>-dirRel</i>	0.8543
<i>+relVec</i>	0.7392
<i>-relVec</i>	0.7673

表 3 に、想起の方向性に関する分類問題の結果 (*Accuracy*) を示す。表において、アスタリスク (*) が付与された値は、属性をすべて用いる場合 (*All*) を上回る結果を示すが、その差は小さく、統計的に有意な差も存在しなかった。

この表に示される結果を想起の強さに対する回帰問題の場合と対比すると、以下のようなことが特徴的である。

- 対称性のある意味的類似度・関連性の貢献度は比較的低い のに対し、非対称性を考慮した属性 (*posSem*, *lexNW*, *dirRel*) の有効性が確認された。

*8 <http://scikit-learn.org/>

*9 <http://chainer.org/>

*10 回帰問題は NN 手法、分類問題は RF 手法が優れるという結果となった。NN 手法に関してはさらに改善の余地があるが、RF 手法は回帰問題に比較的低いという傾向を裏付ける結果である。

- 中でも *lexNW* の単独での有効性は予想を相当に上回っており、関係の方向性を定める上で、語彙資源が提供するグラフ構造における概念ノードのグラフ理論的な重要性の差が重要な手がかりとなる。

一方で、想起関係の方向性の判定というカテゴリカルな問題においても意味的な関係ベクトル (*relVec*) の有効性が顕著に高いことが確かめられた。

表 4 は、属性を全て用いる場合 (*All*) における評価指標 Precision/Recall/F1 の値を関係の方向性ごとに示す。outbound については良好な傾向であるが、inbound の場合の結果はこれに劣る。これは、評価データのもとになった自由連想のデータの特性に影響されているのではないかと想定される。

表 4: 各方向性ごとの評価指標.

Directionality	Precision	Recall	F1
outbound	0.8196	0.9338	0.8730
inbound	0.8788	0.7588	0.8144
bidirectional	0.9586	0.9973	0.9775
no-evocation	0.9510	0.0683	0.1274

5.3 検討: 意味関係ベクトルの種別

以上の結果から、分散表現ベクトルの差分による意味関係ベクトルの有効性が強く示唆される。ただし、これまでに示した結果は、AutoExtend 手法により導出された語彙化概念に対する分散表現ベクトルに基づくもののみであった。そこで、意味関係を表すベクトルとして、W2V ベクトルに基づくもの、AutoExtend 手法により導出された語義ベクトルを用いた場合を評価した。その結果を表 5 に示す。

表 5: 意味関係ベクトルの種別による比較.

Relational vector type	r	ρ	Accuracy
<i>relVec(synset)</i>	0.4391	0.4000	0.8548
<i>relVec(W2V)</i>	0.4551	0.4158	0.7582
<i>relVec(lexeme)</i>	0.4267	0.3880	0.7400

この結果からまず言えることは、想起の方向性の決定においては、語彙化概念 (*synset*) に基づく意味関係ベクトルが他を圧倒していることであり、このようなカテゴリカルな問題において、抽象度の高い表現が有用である可能性が示唆される。一方で、想起の強さの予測においては、W2V に基づく意味関係ベクトルの結果が他より良い。これは、そもそも人間による想起関係の評定時において、単語に影響される評定が行われていた可能性を示唆していると考えられる。また、ある意味で両者の中間に位置する語義 (*lexeme*) ベクトルに基づく意味関係ベクトルは、双方の問題において劣った結果となっている。この原因はさらに詳しく調べる必要があるが、おそらくは、語義が単語と概念の中間層に位置するという AutoExtend 手法における自己符号化器の適用にも関係している可能性がある。

6. おわりに

本論文では、概念間の想起関係という漠然性と多様性が高い意味関係を対象とし、教師付き機械学習の枠組みにおいて、その強さと方向性の予測を行うための手法を提案し、実験的

に評価した。その結果、筆者らによる最近の結果 [13] と比べて顕著に良い結果 (強さの予測: $r \approx 0.44$, 方向性の決定: *Accuracy* ≈ 0.85) が得られた。この違いは、分散表現ベクトルの差分に基づく意味関係ベクトルの利用の有無によることから、分散表現の基本的有用性、および、その差分により意味関係を捉えることの有効性が確認された。

今回報告した結果は、指標の数値的にはまだ十分ではないものの、問題の困難さを考えれば有望な結果である。逆に言えば、さらに精度を向上させるためにはブレークスルーが必要である。非対称的な意味的類似度を対象とする研究 [7, 6] において行われてきたように、コーパスから得られる情報を用いて属性包含に関する扱いを精緻化するほか、語彙知識に潜在する意味的・概念的な関係の利用を高度化する必要がある。後者については、大規模な意味ネットワーク上における意味経路として想起関係を捉える試みを行ってきており [12], このような検討から得られる知見や属性を追加していきたいと考えている。

謝辞

本研究は JSPS 科研費 #26540144 の助成を受けた。

参考文献

- [1] Barthélemy, M. 2004. Betweenness centrality in large complex networks. *European Physical Journal B*, Vol.38, pp.163–168.
- [2] Boyd-Graber, J., et al. 2006. Adding dense, weighted, connections to WordNet. *Proc. of the Third International WordNet Conference*, pp.29–36.
- [3] Budanitsky, A. and Hirst, G. 2006. Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics*, Vol.32, No.1, pp.13–47.
- [4] Cramer, I. 2008. How well do semantic relatedness measures perform? A meta-study. *Proc. of STEP 2008*, pp.59–70.
- [5] De Deyne, S., et al. 2015. Associative strength and semantic activation in the mental lexicon: evidence from continued word associations. *Proc. of the 35th Annual Conference of the Cognitive Science Society*, pp.2142–2147.
- [6] Gawron, J. M. 2014. Improving sparse word similarity method with asymmetric measures. *Proc. of ACL 2014 (Short Papers)*, pp.296–301.
- [7] Kotlerman, L., et al. 2010. Directional distributional similarity for lexical inference. *Natural Language Engineering*, 16 (4): 359–389.
- [8] Ma, X. 2013. Evocation: analyzing and propagating a semantic link based on free word association. *Language Resources and Evaluation*, Volume 47 Issue 3, pp.819–837.
- [9] Mikolov, T., et al. 2013. Distributed Representations of Words and Phrases and their Compositionality. *Proc. of NIPS 2013*.
- [10] Miller, G.A. and Fellbaum, C. 2007. WordNet then and now. *Language Resources and Evaluation*, Vol.41, No.2, pp.209–214.
- [11] Rothe, S., and Schütze, H. 2015. AutoExtend: Extending word embeddings to embeddings from synsets and lezemes. *Proc. of ACL 2015*, pp.1793–1803.
- [12] 林 良彦. 2015. 意味ネットワークの探索と系列パターンマイニングによる想起の類型化. 2015 年度人工知能学会全国大会 2F5-OS-01b-3.
- [13] 林 良彦. 2016. 語彙知識と分散表現を用いた概念間の想起関係の予測. 言語処理学会第 22 回年次大会 D3-1.