

動詞含意関係データベースの自動拡張

橋本 力*† 鳥澤 健太郎* 黒橋 禎夫†§ 藤田 篤‡* 黒田 航*
 ステイン・デ・サーガ* 村田 真樹* 風間 淳一*

* 情報通信研究機構 MASTAR プロジェクト言語基盤グループ § 情報通信研究機構知識処理グループ

† 京都大学情報学研究科 ‡ 公立はこだて未来大学システム情報科学部

1 はじめに

テキスト含意認識は質問応答等の性能向上に寄与する重要な技術だが [3]、その成否は含意知識を効率よく集積する技術にかかっている。本研究では、既存の動詞含意関係データベースを推移律で拡張する手法を提案する。動詞含意とは含意関係が成立している動詞ペアを指す。動詞 1 が動詞 2 を含意するとは、動詞 1 が成立するなら、動詞 2 も成立しているということの意味する。例えば「試乗する」は「運転する」を、「挑戦する」は「チャレンジする」を、「チンする」は「加熱する」を含意する。以下、動詞含意関係を「試乗する→運転する」、「チンする→加熱する」のように表す。

推移律により動詞含意「A→B」と「B→C」から「A→C」が得られるが、単純に適用すると、妥当な「A→B」と「B→C」からも、含意関係にあるとは言い難いペアも生成される。これは、論理的に真ではないが常識に照らして尤もらしい推論的知識や、多義な動詞により引き起こされる。我々は、分布類似度と含意段数の制限により、新規獲得動詞含意の中から適格なものだけを抽出することを試みた。本稿ではその適合率と、元のデータベースにアノテーション無しで追加可能な新規獲得動詞含意数について報告する。

2 関連研究

従来の動詞含意獲得研究はほとんどが教師無し学習によるもので、分布類似度を用いたもの [6, 12, 11, 4] やコンパラブルコーパスを用いたもの [9, 1] 等がある。

分布類似度を用いる手法の大枠は、構文解析されたコーパスから、動詞とその項を大量に収集して、項を素性として動詞間の類似度を計る、というものである。この手法は、大規模なコーパスを用意すれば動詞含意を大量に獲得することができるが、高い精度は期待できず、利用の際は徹底的な人手チェックが必要となる。

コンパラブルコーパスを用いる手法は、同じ出来事に対する複数の異なる記述 (例えば、同じ事件に対す

る複数の新聞記事) をコンパラブルコーパスとして収集し、そこから含意 (言い換え) 知識を獲得する。この手法はコンパラブルコーパスが大量に収集できることを前提としているが、現実には、そのようなドメインはそれほど多くなく、バラエティに富んだ含意知識を大量に獲得するのは困難である。

一方、我々は、高精度と高カバー率の両立を狙い、人手で整備された動詞含意関係データベースを自動拡張する手法を採用した。従来手法が教師無し学習によるものだったのに比べ、本研究の手法は人手で整備されたデータを利用しているという意味で、教師あり学習による含意知識獲得といえる。近年、意味情報が付与された動詞関係の語彙資源 ([2, 5] 等) が数多く世に出回っていることを踏まえれば、人手で整備された語彙資源を出発点として、動詞の意味的知識を大規模かつ高精度に獲得するという手法は実現可能であり、深く掘り下げる価値のある技術であると考えられる。

本研究と類似した試みは動詞含意獲得以外の領域で既に存在する。[10, 7, 13]。[10] は WordNet を、[7, 13] は Wikipedia から得られた上位下位データを自動拡張する手法を提案している。

3 動詞含意関係データベース

本研究で自動拡張の対象とするのは、ALAGIN 動詞含意関係データベース Version 1.1.1 (以下、VEDB) である。VEDB は情報通信研究機構で構築され、高度言語情報融合フォーラム ALAGIN¹ から配布されている。VEDB は約 23,926 の動詞含意の正例と約 38,627 の負例² から構成されている。本研究では負例は使用せず、正例のみを自動拡張する。

VEDB は、[4] の動詞含意獲得実験で用いられた係り受け解析済み Web データに対して [12, 4] のいずれかの手法を適用して得られた動詞含意と、情報通信研

¹<http://www.alagin.jp>

²動詞含意規則の負例とは、例えば「叫ぶ→驚く」等のように、含意関係が成り立たない動詞ペアのことである。

究機構内にある別の解析済み Web データに対して [11] を適用して得られた動詞含意を手手でアノテーションすることで構築された。[12] と [4] で得られた動詞含意はそれぞれスコア上位 50,000 までを、[11] で得られた動詞含意はスコア上位 37,000 までをアノテーションした。アノテーションでは、正解判定だけでなく、一部の動詞含意に対しては語形の修正も行った。例えば「貼付チョウフする」を「貼付する」に、「ランクインする」を「ランクインする」に修正した。

4 提案手法

本研究の VEDB 自動拡張手法は、推移律による自動拡張と、その結果得られた動詞含意のフィルタリングの 2 段階から成る。論理学の世界における推移律は、妥当な「 $A \rightarrow B$ 」と「 $B \rightarrow C$ 」から妥当な「 $A \rightarrow C$ 」を常に生成するが、後述するように、自然言語（あるいは自然言語処理）の世界では、「結婚する→デートする」のような plausible inference³[8] や、多義な動詞も対象に含めるため、推移律により常に妥当な動詞含意が得られるという保証はない。推移律による拡張の後にフィルタリングが必要なのはこの理由による。

4.1 推移律による VEDB 自動拡張

VEDB に推移律を適用して新たな動詞含意を得た⁴。その際、含意段数を 20 までに制限した。含意段数は「 $A \rightarrow B$ 」なら 1、「 $A \rightarrow B \rightarrow C$ 」なら 2 である。結果として、VEDB の正例 23,926 ペアから 408,228 ペアが推移律により新たに得られた。

これら新規獲得ペアには含意が成立しているとは考えにくいものが含まれている。この現象を理解するため、以下で、推移律拡張により妥当な動詞含意が得られる場合とそうでない場合を例とともに説明する。

VEDB に妥当な動詞含意「再婚する→離婚する」と「離婚する→結婚する」がある場合を考える。推移律拡張によりこれらから「再婚する→結婚する」という妥当な動詞含意が得られる。これを図で表すと図 1(a) のようになる。最小の円が「再婚」で、最大の円が「結婚」である。円の包含関係が対応する動詞の含意関係を表す。このケースは論理学における含意に近く、推移律により妥当な動詞含意が得られる。

自然言語処理では「結婚する→デートする」や「デートする→恋する」のような論理的に真ではないが plausible な知識も重要である。これらに推移律を適用すると「結婚する→恋する」という plausible な動詞含意

が生成される。図で表すと図 1(b) のようになる。円の間に完全な包含関係はないが、十分な重なりがある。

「ネットサーフィンする→楽しむ」と「楽しむ→微笑む」は各々 plausible な動詞含意として容認可能だが、これに推移律を適用して得られる「ネットサーフィンする→微笑む」はもはや plausible とは言い難い。これを図示すれば図 1(c) のようになる。そしてこの現象が、推移律拡張により動詞含意として不適格なペアが生成される原因の一つである。

もう一つの原因は動詞の多義性にある。今、VEDB に「徹底討議する→煮詰める」と「煮詰める→煮る」という動詞含意があるとする。「煮詰める」には「十分に議論、検討する」という意味と「水分が少なくなるまで煮る」という 2 つの意味があるため、上記 2 つの動詞含意に単純に推移律を適用すると、「徹底討議する→煮る」という不適格な動詞含意が生成される。

4.2 推移律拡張結果のフィルタリング

本研究では、次の 2 つの方法により、推移律拡張結果中の不適格な動詞含意のフィルタリングを試みる。

1. 含意段数による制限
2. 分布類似度による動詞含意自動獲得結果 [6, 12, 11, 4] 上位 N 位内の有無による制限

1. に関して、推移律拡張結果を調べた結果、含意段数が多いほど誤りが多かった (表 1)。そこで、推移律拡張結果の動詞含意を含意段数で制限する。

2. に関して、推移律拡張で得られた動詞含意のうち、妥当なものはそうでないものに比べて、動詞含意を構成する動詞ペア間 (例えば「再婚する→離婚する→結婚する」の「再婚する→結婚する」) の分布類似度が高いと予想される。そこで、推移律拡張結果の動詞含意を、分布類似度に基づく手法で得た動詞含意リストのスコア上位 N 位に含まれているものだけに制限する。分布類似度に基づく手法として Lin [6]、Precision [12]、BInc [11]、Score [4] の 4 つを採用し、これらを [4] の実験で用いられた解析済み Web データに適用することでスコア付き動詞含意リストを得た。この動詞含意獲得では約 52,000 の動詞を対象とした。

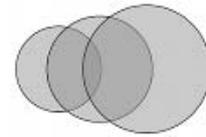
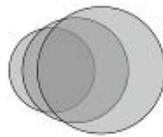
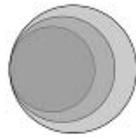
5 評価実験

まず、§4.1 で述べた推移律拡張結果 408,228 ペアから 100 ペアの動詞含意をサンプリングしたところ、適合率は 0.30 だった。

次に、含意段数のみによるフィルタリングで得られたペア数とその適合率を調べた。含意段数として 2、

³以下ではこれらも動詞含意と呼ぶ。

⁴循環する推移律過程は無視した。



(a)「再婚する→離婚する→結婚する」 (b)「結婚する→デートする→恋する」 (c)「ネットサーフィンする→楽しむ→微笑む」

図 1: 推移律拡張で妥当な動詞含意が得られる場合とそうでない場合

表 1: 含意段数 N の適合率とペア数

	段数 2	段数 3	段数 4	段数 5	段数 10
適合率	0.70 (35/50)	0.46 (23/50)	0.44 (22/50)	0.32 (16/50)	0.12 (6/50)
合計ペア数	25,819	27,589	27,343	30,036	24,284

3、4、5、10 を試みた。その結果を表 1 に示す。それぞれの段数で得られた動詞含意群から 50 ずつサンプリングしたところ、0.70、0.46、0.44 と段数が多くなるにつれて適合率が下がることがわかった。

分布類似度での自動獲得結果上位 N 位中の有無による制限を加えた結果は表 2、3 の通りである。N として 50 万と 5,000 万を試みた。表の最下段の合計ペア数は、分布類似度による 4 つの動詞含意獲得手法それぞれで得られたペアの総和から重複を除いたものである。これらの結果からも、段数の増加とともに適合率が下がるという傾向が見て取れる。さらに、分布類似度に基づく動詞含意獲得結果上位 N 位を用いた制限を加えることで、含意段数による制限のみの場合に比べて、適合率が向上することがわかった。

段数 2 から 5 で、いずれかの動詞含意獲得結果上位 5,000 万にあったもののうち、適格なもの例を o-1 から o-4 に、不適格なもの例を x-1 から x-4 に挙げる。

- o-1) 「一煮立ちする→煮る→加熱する」
- o-2) 「伝言する→連絡する→知らせる→告げる」
- o-3) 「侵略する→侵攻する→出兵する→派兵する→派遣する」
- o-4) 「摘出手術する→切除する→手術する→治療する→発病する→煩う」
- x-1) 「巡回する→歩き回る→ぶらつく」
- x-2) 「暴走する→走り回る→ランニングする→ジョギングする」
- x-3) 「街頭演説する→述べる→話す→喋る→おしゃべりする」
- x-4) 「ライブ演奏する→弾く→演奏する→披露する→発表する→表明する」

x-1 と x-3 は推移により出発点の動詞の意味からの乖離が大きくなったため、x-2 と x-4 は各々「走り回る⁵⁾

⁵⁾ 乗り物等で走る意味と人間の足で走る意味。

と「発表する⁶⁾」が持つ多義性のため不適格となった。

本研究はそもそも、既存の語彙資源を自動拡張することが目的であり、追加アノテーション無しでどのくらいデータを増やすことが出来るのかが最も重要な評価指標である。本研究では、適合率 0.90 以上の場合にアノテーション無しで追加可能と判断する。その場合、表 1 にある通り、含意段数の制限のみではアノテーション無しでデータを追加することはできない。分布類似度での自動獲得結果による制限を加えた場合は、含意段数 2 かつ BInc 上位 50 万、5,000 万による制限を適用した場合に適合率が 0.90 となった (表 2、3)。その条件で追加できる動詞含意は、各々 158 ペア、4,056 ペアである。適合率 0.80 以上で、50 ペア以上獲得できた場合を対象とすると、追加できる動詞含意は各々 14,737 ペア、19,989 ペアである。

6 おわりに

本研究では、VEDB を推移律を利用して自動拡張する方法を提案した。VEDB には、論理的に真ではないが plausible な推論的知識や、多義な動詞から成る動詞含意も含まれるため、単純に推移律を適用すると、不適格な動詞含意が生成される。そこで本研究では、含意段数と、分布類似度に基づく動詞含意自動獲得結果による制限を加えることで、適格な動詞含意のみをフィルタリングすることを試みた。その結果、元のデータベースにアノテーション無しで追加可能なデータを得ることに成功した。

今後、フィルタリングの高精度化を狙い、推移律過程の全動詞で共有されている格要素に着目した手法を試みる。この手法は、妥当な動詞含意を生成する推移律過程 (例えば「再婚する→離婚する→結婚する」) はそうでないもの (例えば「徹底討議する→煮詰める→煮る」) に比べて、過程中の全動詞が共通して取り

⁶⁾ 技能等を披露する意味と意見等を公表する意味。

表 2: 分布類似度による動詞含意獲得結果上位 50 万内にあるものの適合率とペア数

	段数 2	段数 3	段数 4	段数 5	段数 10
Lin	0.80 (40/50) 64	0.73 (19/26) 26	0.89 (8/9) 9	1 (1/1) 1	0 (0/2) 2
BInc	0.90 (45/50) 158	0.81 (39/48) 48	0.65 (13/20) 20	0.75 (3/4) 4	1 (1/1) 1
Precision	0.86 (43/50) 5,347	0.82 (41/50) 2,401	0.60 (30/50) 1,349	0.30 (15/50) 1,002	0.32 (16/50) 380
Score	0.86 (43/50) 6,767	0.74 (37/50) 3,048	0.64 (32/50) 1,483	0.50 (25/50) 1,145	0.30 (15/50) 566
合計ペア数 (重複無し)	9,476	4,539	2,446	1,887	836

表 3: 分布類似度による動詞含意獲得結果上位 5,000 万内にあるものの適合率とペア数

	段数 2	段数 3	段数 4	段数 5	段数 10
Lin	0.74 (37/50) 3,031	0.54 (27/50) 1,721	0.48 (24/50) 1,036	0.42 (21/50) 742	0.22 (11/50) 305
BInc	0.90 (45/50) 4,056	0.72 (36/50) 2,104	0.56 (28/50) 1,276	0.52 (26/50) 918	0.26 (13/50) 383
Precision	0.84 (42/50) 15,933	0.62 (31/50) 13,105	0.46 (23/50) 10,900	0.32 (16/50) 10,542	0.20 (10/50) 6,383
Score	0.76 (38/50) 22,123	0.56 (28/50) 20,790	0.36 (18/50) 18,130	0.32 (16/50) 18,555	0.10 (5/50) 12,957
合計ペア数 (重複無し)	22,577	21,725	19,471	20,075	14,206

うる格要素の数が多く、という仮説に基づく。

本研究が目指したのは、人手整備された語彙資源を出発点として、高品質でカバー率の高い語彙資源を得る技術の開発である。本稿では、推移律を利用した動詞含意関係データベースの拡張について報告した。同様の手法は動詞含意以外の語彙資源、例えば異表記データや上位下位データ、類義語データ等にも適用可能であると考えられる。多種多様な語彙資源の整備が着々と進行している現状を踏まえると、それらを出発点として高品質/高カバー率の語彙資源を獲得する技術は、今後の重要な研究課題であると考えられる。

参考文献

- [1] Regina Barzilay and Lillian Lee. Learning to paraphrase: An unsupervised approach using multiple-sequence alignment. *HLT-NAACL*, pp. 16–23, 2003.
- [2] Timothy Chklovski and Patrick Pantel. Verbocean: Mining the web for fine-grained semantic verb relations. *EMNLP*, pp. 33–40, 2004.
- [3] Sanda Harabagiu and Andrew Hickl. Methods for using textual entailment in open-domain question answering. *COLING/ACL*, pp. 905–912, 2006.
- [4] Chikara Hashimoto, Kentaro Torisawa, Kow Kuroda, Masaki Murata, and Jun'ichi Kazama. Large-scale verb entailment acquisition from the web. *EMNLP*, pp. 1172–1181, 2009.
- [5] Karin Kipper-Schuler. *VerbNet: A broad-coverage, comprehensive verb lexicon*. PhD thesis, Computer and Information Science Dept., University of Pennsylvania, 2005.
- [6] Dekang Lin. Automatic retrieval and clustering of similar words. *COLING/ACL*, pp. 768–774, 1998.
- [7] Jong-Hoon Oh, Kiyotaka Uchimoto, and Kentaro Torisawa. Bilingual co-training for monolingual hyponymy-relation acquisition. *ACL*, pp. 432–440, 2009.
- [8] Patrick Pantel, Rahul Bhagat, Bonaventura Coppola, Timothy Chklovski, and Eduard Hovy. ISP: Learning inferential selectional preferences. *NAACL-HLT*, pp. 564–571, 2007.
- [9] Yusuke Shinyama, Satoshi Sekine, and Kiyoshi Sudo. Automatic paraphrase acquisition from news articles. *HLT*, pp. 313–318, 2002.
- [10] Rion Snow, Daniel Jurafsky, and Andrew Y. Ng. Semantic taxonomy induction from heterogenous evidence. *COLING/ACL*, pp. 801–808, 2006.
- [11] Idan Szpektor and Ido Dagan. Learning entailment rules for unary template. *COLING*, pp. 849–856, 2008.
- [12] Julie Weeds and David Weir. A general framework for distributional similarity. *EMNLP*, pp. 81–88, 2003.
- [13] Ichiro Yamada, Kentaro Torisawa, Jun'ichi Kazama, Kow Kuroda, Masaki Murata, Stijn De Saeger, Francis Bond, and Asuka Sumida. Hypernym discovery based on distributional similarity and hierarchical structures. *EMNLP*, pp. 929–937, 2009.