

質問応答システムのための含意パターンペアの生成

川田 拓也 Kloetzer Julien 鳥澤 健太郎 橋本 力

独立行政法人 情報通信研究機構 情報分析研究室

1 はじめに

近年、自然言語処理を利用した質問応答技術の開発が盛んに進められており、我々の研究グループにおいても、2014年度秋に対災害 SNS 情報分析システム DISAANA[9]^{*1}が公開され、2014年度末には情報分析システム WISDOM X[8]を公開する予定である。いずれのシステムも主として事実を問う factoid 型質問応答を主要な機能の一つとして有しており、その factoid 型の質問応答を支える技術が含意の認識技術である。例えば、「肺がんを引き起こすのは何?」という質問に対して、<A が B をもたらす>が<A が B を引き起こす>を含意する（言い換えられる）という知識を持っていれば、「喫煙は肺がんをもたらします」という、質問とは異なる表現から回答を得ることが可能となるため、回答を得られる質問の数が増えることが期待される。また、一つの質問に対して、より多くの回答がヒットすることで、意外なものも含む多様な回答が得られることが期待される。そのためには、このような含意関係にある言語パターンペアを効率的かつ大量に収集する必要がある。

本稿では項が一つもない動詞の含意ペアを種に、単純な規則を適用することで、項が1つ以上のより複雑な言語パターンの含意ペアを生成し、パターンの数を拡張する手法について報告する。また、本提案手法が、単純なヒューリスティクスのみで、大量の精度の高い含意ペアを効率的に生成できることを示す。

次節では本稿が扱う言語パターンの定義と、含意の考え方について述べる。次に、具体的な含意パターンペアの生成手法について述べ、次に生成した含意パターンペアの評価を行う。最後に本稿をまとめる。

2 パターン間の含意関係

我々が開発している質問応答システムは、述語とそれに係る1つ以上の名詞句からなる言語パターンがインデックス化され、検索時に用いられている。例えば、<A が B をもたらす>や<A が B を引き起こす>というようなパターンである。そのため、我々が対象とする含意のパターンペアもそのような言語パターンのペアを想定している。本稿では、上記のように述語とそれに係る項が2

つの言語パターンをバイナリパターンと呼ぶこととする。例えば<A が B をもたらす>というようなパターンである。また、項が1つの言語パターンをユナリパターンと呼ぶこととする。例えば<A をもたらす>といったパターンである。また、本稿における含意のパターンペアは、<P>→<Q>で表わされ、前件であるパターン P が成立する時、後件であるパターン Q も同時に成立するような関係を指す。

これまでパターン間の含意関係の構築を進めてきたが、主に、機械学習をベースとする自動獲得によるものであった [6, 5, 4]。しかし機械学習による獲得結果のみを利用しようとする、具体的にどのようなペアが獲得できるかを予測することは難しく、精度も保証はできない。本手法は、ある程度人手で構築したデータを種に、必要な含意ペアを効率よく獲得できる利点がある。

3 含意ペアの拡張

3.1 ユナリパターンペアへの拡張

本節では、項がない動詞ペアからユナリパターンペア、さらにはバイナリパターンペアへの拡張方法について説明する。なお、図1において、本節で説明する拡張手法の概要を示した。

本手法を適用して、含意パターンペアを生成するには、最初に種となる動詞の含意ペアを用意する必要がある。そのため、我々は高度言語情報融合フォーラム^{*2}にて公開されている動詞含意データを利用した。これは「もたらす→引き起こす」といった含意ペアが人手で構築されたデータで、52,689 ペア収録されている。まずはこの動詞の含意ペアに対して助詞を付与することでユナリパターン化を施した。付与する助詞は「が」「に」「を」「で」のいずれかに限定してペアを作り、それぞれペア同士の文脈類似度 [2] が最も高いペアを選択することでより意味の近いペアとなるようにした。例えば<A を引き起こす>→<A が発生する>といったペアである。これをさらに人手でチェックした結果、構築された正例のユナリパターンペアは51,589 ペアとなった (図1における VEDB)。

さらに、Sano ら [7] の Phased Predicate Template Taxonomy (PPTT) や Hashimoto らの述語の活性不活

^{*1} <http://disaana.jp/>

^{*2} <https://alaginrc.nict.go.jp/>

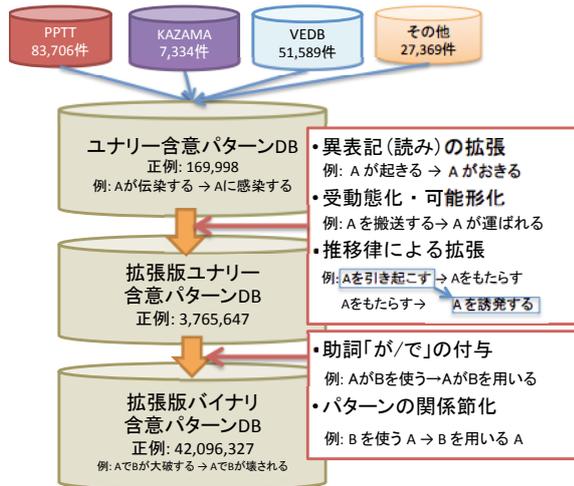


図1 パターン拡張の概要

性 [1] を利用して、ユニナリパターンの含意ペアも作成した。PPTT は述語の新しい意味機能分類で、時間的前後関係などのユニナリパターン間の意味的關係によって分類した体系である。また、活性不活性は述語がその係り元の名詞の機能や効果を発揮させるか否かに基づいて、ユニナリパターンを分類したものであるが、PPTT は活性不活性の分類も利用して構築されている。PPTT 由来の含意ペアデータを構築するにあたって、PPTT の分類の中で「一方のパターンが起きているのであればもう一方のパターンも同時に起きている」という関係 (PPTT では Present+ と呼ばれる) などの述語クラスを利用した。そこで得られたペアのうち、文脈類似度の上位のペアを手手でチェックを行い、正例が 83,706 ペアから成るデータを構築した (図 1 における PPTT)。

また、機械的に生成した含意ペアも利用した。Kazama による文脈類似度 [3] が付与された述語対から係り元の助詞が「を」「に」「で」の述語を選び、ユニナリパターンとした (図 1 における KAZAMA)。また、<A をする/持つ/おこなう>など、意味が曖昧なパターンは除外した。この手法で得られたペアを手手で判定し、正例と判定された 7,334 ペアを用いた。その他、我々がこれまで構築してきたいくつかのデータ (例えば、人手で因果関係にあるようなパターンを収集したデータなど) を基に 27,369 ペアの正例を種として用いた。

次は、得られたユニナリパターンを述語の表記揺れの対応や語尾の変形によりパターンを拡張した。表記揺れの対応については、単純に、漢字で書かれた述語を仮名に変換したパターンを作り、新たなパターンとして加えた。また、<A が起きる> → <A がおきる> というような、同じ述語の「漢字 → 仮名」「仮名 → 漢字」というペアも同時に生成し、追加した。「仮名 → 漢字」のペアは厳密には「含意」ではないが、質問応答における「言い換え」という

文脈においてはそのようなペアも必要であると判断して、含意とみなしている。

上記の手法で得られたユニナリパターンの述語は原則として基本形である。質問応答システムという文脈においては、質問や回答が受動態や可能形で現れることが少なくないため、受動態や可能形パターンも生成して拡張する。しかし、特に受動態は格交替が生じるため、述語の語尾変化だけでなく、パターンに含まれる格を変化させる必要がある。そこで、ユニナリパターンを以下の規則で可能形、受動態に変換してパターンを拡張した。可能形への変換は、格助詞が「を」「で」のパターンに限定した。この制約は助詞が「は」や「が」の場合、単純に可能形とすると、その項が対格に交替する可能性があることによる (例えば、<A が食べる> → <A が食べられる> としてしまうと A が対格としても解釈可能となる)。可能形への変換については、動詞の場合は助動詞「れる/られる」を付加し、「サ変名詞+する」の場合は、「サ変名詞+できる」と変換する単純な規則とした。受動態への変換は、助詞が「を」のパターンのみを対象とし、その場合は助詞を「が」に変換し、述語に対して、助動詞「られる」(サ変動詞の場合は「される」) を付加する。例えば、<A を搬送する> → <A を運ぶ> というペアに対して、可能形と受動態を作り、拡張すると以下のようなパターンを生成できるようになる。

A を搬送する → A を運ぶ	A を搬送する → A が運ばれる
A を搬送する → A を運べる	A を搬送できる → A を運ぶ
A を搬送できる → A が運ばれる	A を搬送できる → A を運べる
A が搬送される → A を運ぶ	A が搬送される → A が運ばれる
A が搬送される → A を運べる	

さらに、以下のような述語が同じ可能形 (受動態) と基本形のペアも同時に生成する。

A を搬送する → A を搬送できる	A を搬送できる → A を搬送する
A を搬送する → A が搬送される	A が搬送される → A を搬送する

上記のように生成することで、ユニナリパターンペアのバリエーションを容易に拡張できるようになる。

次に「推移律」による拡張を行う。二組の含意ペア $P \rightarrow Q$ と $Q \rightarrow R$ が成立する時、推移律によって $P \rightarrow R$ も成立する。この論理関係を利用すれば、二組の含意ペアから新たな含意ペアを生成することが可能となる。例えば <A を引き起こす> → <A をもたらし> というペアと、<A をもたらし> → <A を誘発する> というペアがある時、この二組のペアから、新たに <A をもたらし> → <A を誘発する> というペアを生成する。ただし、人手で判定、チェックした種となるデータには、その過程で負例と判定されたペアもあるので、推移律で生成したペアが人間の判断ですでに負例としてアノテーションされている場合は登録していない。なお、同音異義語が誤って登録されないよう、同じ読みの述語の漢字と平仮名のペアについては推移律を適用していない (<A が空く> → <A があく>、<A

があく>→<A が開く>から<A が空く>→<A が開く>を生成しないようにする)。現状では、推移律を適用する前の含意パターンペア数が 901,232 ペアに対して、推移律による拡張を行うことで、新たに 2,864,415 ペアが生成されており、約 3 倍に拡張されている。

上記手法で生成されたユナリパターンペアは数値としてはスコアリングされないが、次のような方法で質問応答システムに利用される。まず、各パターンの Web6 億文書中における頻度を計算しておく。その上で、あるパターンがヒットした場合は、ユナリパターンのデータの由来や性質に応じた、以下の優先順位に従って優先的に言い換えパターンとして利用される。

1. 同値関係 ($P \rightarrow Q$ と $Q \rightarrow P$ が共に成立する関係)
2. 読みや形態 (可能形・受動態) が異なる同述語
3. 上記以外で KAZAMA, PPTT, VEDB を基に構築
4. 推移律によって生成
5. データの由来に関わらず Web6 億文書中の頻度が 10 未満

さらにその中で頻度が高いパターンから順位付けされる。最終的には上位パターンが言い換えパターンとして用いられる。

3.2 バイナリパターンペアへの拡張

質問応答システムにおいてユナリパターンは回答としてヒットしやすい反面、項が一つしかないため、回答精度に過度な期待はできない。従って、我々が構築しているシステムでは、最初に項が 2 つのバイナリパターンペアを元に回答を検索し、それでも回答が見つからなかった場合にユナリパターンペアから回答を検索するような仕組みになっている。そのため新たなバイナリパターンペアを生成することで拡充を図る。

バイナリパターンは<A が B をもたらす>というような主要部が述語のパターンに加え、<B をもたらす A>という一方の項を主要部とする連体節のパターンも存在する。そこでユナリパターンからバイナリパターンへ拡張する際は、主要部が述語のパターンと項のパターンの両方に拡張をする。拡張方法は次の二通りである。第一に、助詞が「が」ではないパターンについて、「A が」を付加することで、主要部が述語のバイナリパターンに拡張する。第二に、パターン末尾に「A」を付加することで、連体節化されたバイナリパターンを生成する。同じように、助詞「で」を付加してバイナリパターンを生成する。具体的には<A を使う>→<A を用いる>からは以下のバイナリパターンペアが生成される。

A が B を使う → A が B を用いる	A が B を使う → A が用いる B
A が B を使う → B を用いる A	A が使う B → A が B を用いる
A が使う B → A が用いる B	A が使う B → B を用いる A
B を使う A → A が B を用いる	B を使う A → A が用いる B
B を使う A → B を用いる A	

表 1 ユナリパターンペアの評価結果

	正例 (精度)	合計
推移律で生成したペア	249 (0.64)	386
推移律以外で生成したペア	94 (0.82)	114
全てのペア	343 (0.68)	500

なお、ユナリで含意が成り立つと判定されたペアに対して、助詞を付加したパターンと連体節化したパターンからなるペアを 400 ペア (「が」と「で」でそれぞれ 200 ペアずつ作成) を作成し、1 ペアにつき 1 人の著者以外の作業員 (計 3 名) が評価した結果、「が」を付加したペアは 179/200(0.90)、「で」を付加したペアは、131/200(0.65) で含意成立とみなされ、全体の精度は 0.78 となった。

さらに、同じユナリパターンを組み合わせたペアを作り、上と同様の操作を行う。すなわち<A を使う>というパターンから<A が B を使う>→<B を使う A>といったペアを生成する。上記の手法によって、最終的にバイナリパターンペアは 42,096,327 ペアが生成された。種となるユナリパターンペアが 169,998 ペアだったので、最終的には約 280 倍の規模のデータが構築されたことになる。

最終的には生成されたペアはヒューリスティクスにより、以下のようにスコア付けされる。

$$A - (A - B) / (f / 1000 + 1)$$

この時、 f は Web6 億文書中に現れた二つのパターンの頻度の合計を表す。また A, B の値は、元となるユナリパターンに応じて以下のように与えられる。

- 同値関係にあるパターン: $A = 0.2, B = -0.2$
- 推移律で拡張したパターン: $A = -0.2, B = -0.8$
- 上記以外のパターン: $A = 0.1, B = -0.3$

4 評価

本節では本手法で得られた含意ペアの評価について述べる。ユナリパターンペア (3,765,647 ペア) については、いずれかのパターンの頻度が 10 以上のペアについてランダムサンプリングによって 500 ペアを抽出し、著者以外の作業員 3 名に依頼して評価を行った。 κ 値 (Fleiss) は 0.46 で中程度の一致となった。作業員自身がパターンの項をいくつか埋めて判定する必要があるため、通常の含意判定と比較すると判定が難しかった可能性がある。評価に当たっては、3 名の判定の多数決を最終的な評価結果とした。図 2 はユナリパターンペアを優先順位の高い順にソートし、上位 n 位の精度をプロットしたものである。全体としては 0.68 程度の精度となったが、表 1 が示すとおり、推移律以外で生成したペアについては 0.8 を超える精度となった。

同様に、生成されたバイナリパターンペア (42,096,327 ペア) から 500 ペアをランダムサンプリングして評価し、そのスコア順にソートしてプロットした結果を図 3 に示

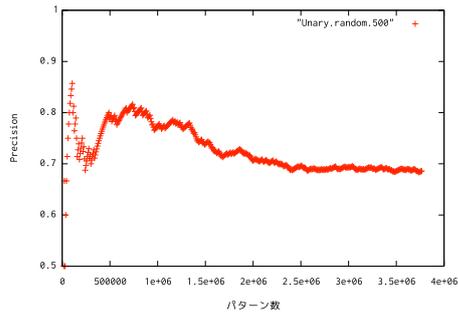


図2 ユナリパターンペアの精度

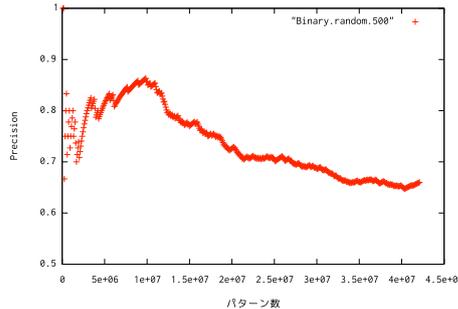


図3 バイナリパターンペアの精度

す。このグラフが示すとおり、全体としては0.66となり、ユナリと同程度の精度となった。

いずれも推移律を用いたペアや、頻度の低いペアが精度を引き下げている原因になっているものの、質問応答システムにおける言い換えという観点から言えば、約4,200万のバイナリペアでは上位30%で約1,200万ペア（約380万のユナリペアでは上位20%で約75万ペア）に関して精度が0.8程度あり、十分実用的なデータであると考えられる。実際のシステムでは、本手法で得られた約4,200万のバイナリペアに加えさらに機械学習を用いて獲得したペアも利用しており、トータルで約1億の含意ペアが用いられている。

実際にシステムを使用して「ビッグデータは何に使うか」という質問を入力すると、「ビジネス」や「マーケティング」といった予想可能な回答に加えて、<Aに役立つ>→<Aに使う>という推移律を基に得られたペアから「輸送計画に役立つ」や「ビッグデータの活用が正確な政治的予測にいかに関係が盛んに報じられる」といった回答が得られ、さらに「信長は何を狙ったか」という質問を入力すると、「天下統一」といった常識的な回答だけではなく、{<AがBを目指す>/<BをめざすA>}→<AがBを狙う>という推移律を基に得られたペアから「絶対王政を目指す」や「自由経済をめざす」といった回答が得られており、本手法で拡張したパターンから多様な回答が抽出されていることが確認できた。

5 おわりに

本稿では、単純な動詞の含意ペアから項が1つ以上の複雑な言語パターン間の含意ペアデータを、簡便な規則を適用することで生成する手法について述べた。この手法により、必要な動詞の含意ペアを用意さえすれば、それに関連する様々なパターンのペアが自動的に生成されるため、機械学習で獲得したペアだけでは網羅できなかったペアを大量に追加できる。システムの性能に関する定量的な評価は今後行う予定ではあるが、本手法の適用後、システムが取得できる回答が大幅に増加したことは確認できており、機械学習による獲得手法に加え、このような単純なヒューリスティクスを用いた方法を用いて含意ペアを拡張することも、質問応答システムにおいては非常に有効であり、今後両者をうまく組み合わせることで、さらなる性能の向上が期待される。今後は推移律で生成したデータを人手で修正するなどして精度を高めると共に、新たなデータを適宜追加することでより網羅性の高い含意パターンペアを構築していく予定である。

参考文献

- [1] Hashimoto, C., Torisawa, K., De Saeger, S., Oh, J.-H. and Kazama, J.: Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 619–630 (2012).
- [2] Hashimoto, C., Torisawa, K., Kuroda, K., De Saeger, S., Murata, M. and Kazama, J.: Large-Scale Verb Entailment Acquisition from the Web, *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 1172–1181 (2009).
- [3] Kazama, J., De Saeger, S., Kuroda, K., Murata, M. and Torisawa, K.: A Bayesian Method for Robust Estimation of Distributional Similarities, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 247–256 (2010).
- [4] クロエツエージュリアン, 鳥澤健太郎, デ・サガステイン, 佐野大樹, 橋本力, 後藤淳: 含意関係パターンペアの大規模獲得, 2013年度情報処理学会関西支部支部大会 (2013).
- [5] Kloetzer, J., De Saeger, S., Torisawa, K., Sano, M., Goto, J., Hashimoto, C. and Oh, J. H.: Supervised Recognition of Entailment Between Patterns, *言語処理学会 第18回年次大会発表論文集*, pp. 431–434 (2012).
- [6] Lin, D. and Pantel, P.: DIRT — Discovery of Inference Rules from Text, *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 323–328 (2001).
- [7] Sano, M., Torisawa, K., Kloetzer, J., Hashimoto, C., Varga, I. and Oh, J.-H.: Million-scale Derivation of Semantic Relations from a Manually Constructed Predicate Taxonomy, *Proceedings of COLING 2014*, pp. 1423–1434 (2014).
- [8] Tanaka, M., De Saeger, S., Ohtake, K., Hashimoto, C., Hijiya, M., Fujii, H. and Torisawa, K.: WISDOM2013: A Large-scale Web Information Analysis System, *The Companion Volume of the Proceedings of IJCNLP 2013: System Demonstrations*, pp. 45–48 (2013).
- [9] 後藤淳, 大竹清敬, De Saeger, S., 橋本力, Kloetzer, J., 川田拓也, 鳥澤健太郎: 質問応答に基づく対災害情報分析システム, *自然言語処理*, Vol. 20, No. 3, pp. 367–404 (2013).