

含意関係を用いたパタン間矛盾関係獲得

Julien Kloetzer 鳥澤健太郎 Stijn De Saeger 佐野大樹 橋本力 呉鍾勲 大竹清敬
 情報通信研究機構 (NICT) 情報分析研究室*

{julien, torisawa, msano, ch, rovellia, kiyonori.ohtake}@nict.go.jp

1 はじめに

文間の矛盾関係認識は、自然言語処理分野において重要なタスクの一つである。[9]では、矛盾関係認識を利用することで、ネット上の不正確な情報やデマを検出する支援となることが指摘された。例えば、「アガリクスは癌に効果的です」といったテキストはWeb上によくあるが、一方で「アガリクスを含む一部の製品に発がんを促進する作用が認められた」と主張する国の研究結果もネット上には存在する。これらのテキスト間の矛盾が自動的に認識できれば、そもそもアガリクスのガンに対する予防、治療効果に疑いの目を向けることが可能となろう。

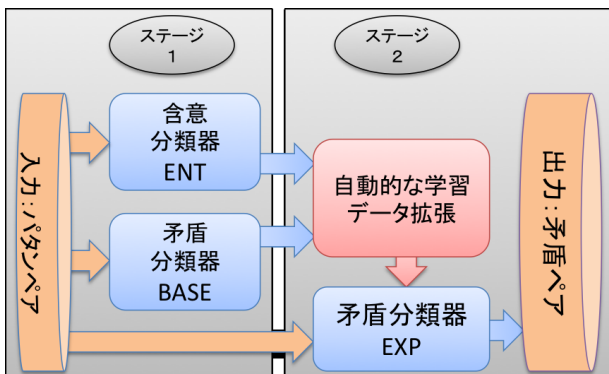


図 1: 提案手法ワークフロー

本稿では(“XがYを予防する”, “XがYを促進する”)のような矛盾関係にある二変数パタンペアを獲得する手法を提案する。この手法は、図1に示すように、ステージ1・2で教師あり学習を用いた分類器を3つ使用するもので、ステージ1では、**BASE**と**ENT**という矛盾認識のための分類器と**ENT**と**含意認識分類器**を作成する。ステージ2では、**BASE**で獲得された矛盾データと**ENT**で獲得された含意データを組み合わせて**BASE**の学習データを自動的に拡張し、**EXP**という新しい矛盾認識分類器を作成する。含意を利用して矛盾の学習データを拡張する点が本手法

の特徴である。**BASE**のみを使用した場合、精度80%で約28.5万件の矛盾ペアを獲得するが、**EXP**を使用した場合では、同じ精度で約75万ペアを獲得する事が可能となる。なお、詳述するスペースはないが、[1]の含意関係獲得手法をベースに Integer Linear Programming を用いた手法も開発したが、提案手法の方が高い精度を得られたことを付記しておく。また、従来の矛盾認識の研究 [4, 6] はパタン間でなく、ドキュメント間や文間の矛盾を認識するものであるが、そうした長い表現間の矛盾も本稿で獲得する矛盾パタンペアを利用することで性能向上が図れると期待される。

2 提案手法

提案手法の新規性は、先述したとおり、パタン間の含意関係と矛盾関係を組み合わせて自動的に新しい矛盾ペアを作成することで学習データを自動拡張することにある。具体的には、パタンペア $\langle p, q \rangle$ が $p \Rightarrow q$ という含意関係を持ち、 $\langle q, r \rangle$ が矛盾ペアである場合、 $\langle p, r \rangle$ は矛盾ペアになるという論理的関係を用いてデータを拡張する。これにより、含意ペアと矛盾ペアから新しい矛盾ペアを獲得する事が出来、学習データの正例を増加させることができる。

問題は、**BASE**の出力する矛盾ペアと**ENT**の含意ペアを組み合わると各分類器の誤りによって誤った矛盾ペアが得られてしまうことであり、結果として最終的に得られる**EXP**の精度を低下させてしまう。これを可能な限り回避することが重要であるが、我々はCDPというスコアを考案し、この問題に対処する。

以下では、まず、本研究のターゲットとなるパタンと各分類器について述べ、ついで上述の学習データ拡張法について詳述する。ただし、スペースの都合上、分類器**ENT**についての詳細は省くが、[8]に詳細が記載されているので参照されたい。

2.1 ターゲットとするパタン

本研究では「 $X_{薬}$ が $Y_{病気}$ を防ぐ」のような、特定の意味クラスに属する変数二つを含むパタンを扱う。

*第3著者が所属している時に本研究が行われた

「薬」や「病気」は変数が属する意味クラスをあらわす。このように意味クラスを用いて変数を埋める名詞に制約をかけることで、パタンの意味的曖昧性を解消することができるようになる [3] (例:「 $X_{日付}$ にある $Y_{イベント}$ 」と「 $X_{国}$ にある $Y_{山}$ 」)。具体的なデータとしては、まず、6億ウェブページのコーパスから「 X [助詞] Y [助詞][動詞]」という形を持つ390万パターンとそのパターンと共起する名詞ペアを抽出し、また、[7]の提案手法を利用して100万個の名詞句をカバーする500個の意味クラスを得た。なお、上の例では「薬」といったラベルのついた意味クラスがあるかのように書いたが、実際には意味クラスのラベルはEM法での隠れ変数に対応するIDとなる。各分類器の入力としては、両方のパターンと共起する名詞ペアが3つ以上ある7.92億パターンペア(セット P_{all})を収集した。これは、含意関係、矛盾関係にあるパターンペアは多数の名詞ペアを共有しやすいという経験則に基づいている。

2.2 BASE : ステージ1の矛盾分類器

ステージ1の矛盾分類器BASEには、[5]が提案した活性・不活性極性を利用した。活性・不活性極性は、「 X を引き起こす」といった一変数のパターンを「活性」「不活性」「中立」に分類する。活性には、変数 X の指す対象の主たる機能、効果、目的、役割、影響が準備あるいは活性化されることを含意するものが該当する(例:「 X を引き起こす」、「 X を使う」、「 X が増える」)。不活性には、 X の指す対象の主たる機能、効果、目的、役割、影響が抑制あるいは不活性化されることを含意するものが該当する(例:「 X を防ぐ」、「 X を捨てる」、「 X が不可能になる」)。中立は、活性でも不活性でもないものである(例:「 X を考える」、「 X に比例する」)。活性・不活性極性が反対となる一変数パタンのペアは矛盾関係になる確率が高くなる [5] ため、矛盾分類器のBASEとEXPの入力データとなる二変数パターンペアを活性・不活性が反対になる一変数パターンペアを部分パターンとして含むものに制限することとした。これは、例えば、「 X を引き起こす」「 X を防ぐ」は、活性・不活性極性がそれぞれ、活性、不活性で反対となるが、「 Y が X を引き起こす」「 Y が X を防ぐ」のようにそうした一変数パターンを部分パターンとして持つ二変数パタンのペアを矛盾分類器の入力とするということである。この入力の制限では、[5]にある自動獲得手法の出力を人手で検証して作成した6,470個の一変数パターンを含む活性・不活性ラベルの辞書を用い、結果として約800万個の二変数パタンのペアを得た。このパターンペアの集合を以下では、 P_{opp} とする。

表 1: BASE の学習データサンプル

二変数パターンペア	ラベル
Y も X が悪い - Y でも X が良い	矛盾
Y も X に向かう - Y も X を出す	矛盾
X に Y を加える - X を Y に入れる	非矛盾
Y も X に来る - Y とは X に行く	非矛盾

活性・不活性極性の効果 活性・不活性極性を用いる効果を調べるため、後述する方法で作成したBASEで(活性・不活性極性が反対となる) P_{opp} のパターンペアと(全てのパターンペア) P_{all} のパターンペアを両方分類し、3名のアノテータの多数決によって矛盾・非矛盾ラベルが付与された評価データにおいてBASEの矛盾認識精度をはかった。(P_{opp} は500ペア、 P_{all} は250ペアの評価データで評価した。) 図2に、その精度を示す。活性・不活性極性を利用することでBASEの精度が向上することが分かる。また、もう一つの方法としてBASEの素性に活性・不活性極性を含めることも試したが、顕著な性能向上は見られなかった。

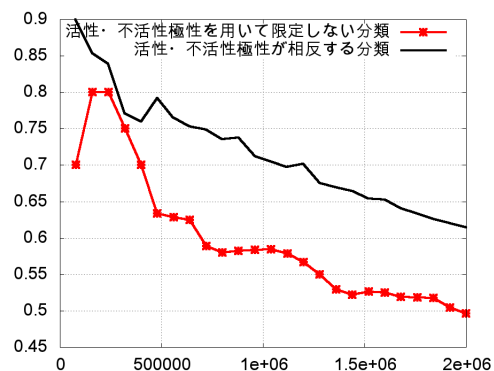


図 2: 活性・不活性極性の効果

学習の詳細 BASEは以下のようにして作成した。まず、学習データとして、人手で約800件の矛盾・非矛盾ラベルが付与された一変数パターンペアからそのペアを含む P_{all} にある二変数パターンペアを収集し、矛盾ペア25.6万件、非矛盾ペア520万件のセットを作成した(表1参照。もとの一変数パターンペアには下線)。性能が最も高くなる学習データの量、矛盾ペア、非矛盾ペアの割合を開発データセットでのトライアルにより決定した。具体的には学習データセットは矛盾ペア12,500件、非矛盾ペア10万件を含む。なお素性として表層的な素性や統語的な素性を利用した([8]を参照されたい)。学習には3次の多項式カーネルを用いてTinySVM¹を利用した。

¹ <http://chasen.org/~taku/software/TinySVM/>

表 2: $\langle p, q, r \rangle$ 例: p が q を含意し、 q が r と矛盾するので p が r と矛盾する

パターン p	パターン q	パターン r	X/Y 例	SVM スコア (p, r)	CDP(p, r)
Y から X が消える	Y から X が無くなる	Y が X に満ちる	怒り/眼	0.3	0.98
Y に X を停止する	Y に X を終える	Y から X を始める	4月/活動	-0.3	0.61
X は Y を示す	X が Y を持つ	X は Y を失う	チーム/自信	0.07	0.45

2.3 EXP : ステージ 2 の矛盾分類器と学習データ拡張方法

次に、BASE と ENT によって得られた矛盾ペアと含意ペアを用いて、ステージ 2 の矛盾パターン分類器 EXP のための学習データを BASE の学習データを拡張することによって得る。この拡張は「もしパターン P (例:「X が Y を引き起こす」) がパターン Q (例:「X が Y を促進する」) を含意し、かつパターン Q がパターン R (例:「X が Y を予防する」) と矛盾するのであれば、パターン P はパターン R と矛盾することとなる」という論理関係に基づき行った。ここで $\langle Q, R \rangle$ をソース矛盾ペアとし、 $\langle P, R \rangle$ を拡張矛盾ペアと呼ぶこととする。このアイデアをもとに BASE の出力 (以下、 C と呼ぶ。具体的には BASE の出力のうち、SVM スコアの上位 5% のペア全てを含む)、と ENT の出力 (以下、 E)。ENT の SVM スコアが 0 以上のペアを全て含む) を組み合わせて BASE の学習データを拡張する。具体的なプロセスを Algorithm 1 に示し、拡張されたペアの例を表 2 に示す。

Algorithm 1 学習データ拡張プロセス。C は BASE の出力上位 5%、E は ENT のスコアが 0 以上の出力を示す

- 1: **procedure** 拡張(C, E)
- 2: 拡張ペアセットの計算 $C' = \{\langle p, r \rangle \mid \exists q : \langle p, q \rangle \in E, \langle q, r \rangle \in C\}$
- 3: C' のペアを CDP によって順位付ける
- 4: $C' \setminus C$ の上位 N ペアを新ポジティブ学習データとして BASE の学習データに追加する
- 5: 異なる学習データサンプルを外す
- 6: **end procedure**

拡張ペアのランキング ここで問題は、本節冒頭でも述べたように、ソース矛盾ペアの誤りが拡張プロセスを経由して、誤りを増加させる可能性があるため、精度の高い拡張ペアを選択する必要があるということである。我々の予備実験では、BASE のスコアに従って高精度と思われるソース矛盾ペアを選択しても、それではわずかな性能改善しか見られなかったため、新たに *Contradiction Derivation Precision* (以下、CDP) というスコアを提案する。CDP ではまず各ソース矛盾ペアの信頼度 (以下 CDP_{sub}) を計算する。この信

頼度は、直感的にはそのソース矛盾ペアをなんらかの含意ペアと組み合わせて、複数の拡張矛盾ペアを得た場合に、それらの拡張矛盾ペアがどの程度正しいかを見るものである。もちろん、拡張矛盾ペアが正しいかどうかは究極的には人の判断によるしかないが、それでは自動化ができないため、拡張矛盾ペアが BASE によって矛盾として認識されるかどうかによって、この信頼度を計算する。この信頼度 $CDP_{sub}(q, r)$ はソース矛盾ペアから拡張されたペアの殆どが BASE によって矛盾ペアとされた場合に高くなるように設計されており、具体的には以下のように定義される。 α という閾値を決定し、拡張によって得られたペアに対する BASE のスコアが α 以上であれば矛盾とし、 α 以下であれば非矛盾とする。

$$CDP_{sub}(q, r) = \frac{|\{\langle p, r \rangle \in Ex(q, r) \mid Sc(p, r) > \alpha\}|}{|Ex(q, r)|}$$

$Ex(q, r)$ はソース矛盾ペア $\langle q, r \rangle$ から、含意ペアとの組み合わせにより拡張されたペアの集合で、 Sc は BASE の SVM スコアである。本稿では BASE の出力の上位 5% ペアのスコアが α 以上になるように $\alpha = 0.46$ にセットした。

これから拡張ペアの CDP を次のように計算する。これは、拡張された矛盾ペア $\langle p, r \rangle$ について、その元となったソース矛盾ペア $\langle q, r \rangle$ は複数個あり得るが、それら全てに対して信頼度 $CDP_{sub}(q, r)$ を計算し、その最大値を値とするものである。

$$CDP(p, r) = \max_{\langle q, r \rangle \in Source(p, r)} CDP_{sub}(q, r)$$

最終的には、BASE と ENT の出力を組み合わせて得た拡張矛盾ペアを CDP によってランキングして上位 N ペアを BASE の学習データに追加し、得られたペアの集合を EXP のための学習データとする。なお、重要なポイントとして、EXP の学習データに追加される拡張矛盾ペアには、BASE のスコアが極めて低く、BASE には矛盾とみなされなかったものも含まれ、こうしたペアは BASE と EXP の出力に大きな差をもたらす、最終的には性能向上に貢献すると思われる。

最終的な学習データの構造 BASE の学習データの負例には、追加された拡張矛盾ペア、つまり、正例が含まれているケースがあるが、そうした場合には負例の

方を削除する。まとめると、学習データ拡張プロセス後に新しい分類器の **EXP** の学習データセットは次の (1)~(3) によって構成されることとなる。(1) **BASE** の学習データのポジティブペア、(2) *CDP* スコアにランキングされた拡張、かつ **BASE** スコアが α 以下であるペアの上位 N ペア、と (3) 2 と異なるペアを抜いた **BASE** の学習データの負例ペア、である。

3 評価

3.1 開発データとテストデータ

開発データ、テストデータとして P_{opp} からランダムサンプリングされた 3,000 ペア (開発データ 1,000 件、テストデータ 2,000 件) を利用する。サンプルはアナテータ 3 名が矛盾・非矛盾かを判定し、最終的なラベルは多数決によって決定した。 κ 値は 0.523 であった。矛盾の定義は [2] に準拠する。

3.2 実験結果

比較手法は次の 3 つである: (1) **PROPOSED** は本稿の提案手法 ($N = 6,000$ は開発データを用いた予備実験により決定された)、(2) **BASE** はステージ 1 の矛盾分類器、(3) **PROP-SCORE** は *CDP* の代わりに SVM スコアを使っている以外は 1 と同じ手法である ($N = 30,000$ は開発データを用いた予備実験により決定された)。テストデータにおける適合率を図 3 に示す。

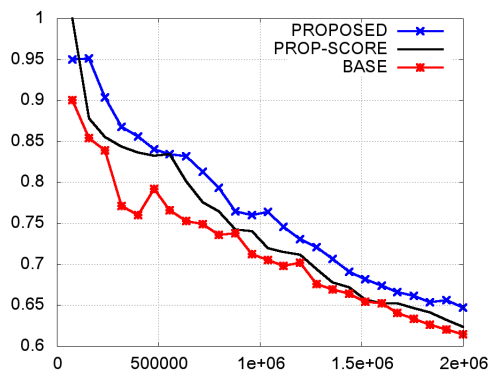


図 3: 比較する全手法の精度

BASE より **PROPOSED** の適合率の方が顕著に高い。**PROPOSED** を使用して P_{opp} の全ペアを分類すると精度 80% で約 75 万矛盾ペアを獲得できた。そのペアから得られた例を表 3 に示す。

BASE では同じ適合率で約 28.5 万ペア、**PROP-SCORE** では約 63.6 万ペアしか獲得できていない。提案手法を利用することで **BASE** の倍以上のペア数を獲得でき、**PROP-SCORE** より高精度で矛盾ペアを獲得できることは、*CDP* の有用性を示すものと考えられる。

表 3: **PROPOSED** によって獲得されたペアとそのラベル (矛盾 (+)、非矛盾 (-))

	パターンペア	X/Y 例
+	Y で X が終わる - Y より X を開始する ランク 228,039	販売/昨日
+	X が Y に勝つ - Y が X に勝つ ランク 258,068	日本/ベトナム
-	X は Y を失う - X には Y はある ランク 474,143	人/興味
+	Y に X を無くす - Y にも X をもつ ランク 522,534	自信/自分
+	Y を X から追い出す - X に Y が残る ランク 697,177	体/疲労
+	Y が X を安心させる - Y が X を裏切る ランク 749,916	僕/彼女

4 おわりに

本稿では大規模に矛盾関係のペアを獲得するための手法を 2 つ提案した: (1) 含意関係と矛盾関係の相互作用を利用することで自動的に学習データを拡張する手法と (2) 活性・不活性の利用による二変数パターン間の矛盾関係獲得。これらの手法を利用して適合率 80% で約 75 万矛盾ペアを獲得する事ができた。

参考文献

- [1] J. Berant, I. Dagan, and J. Goldberger. Global learning of typed entailment rules. In *Proceedings of ACL 2011*, pages 610–619, 2011.
- [2] M.-C. De Marneffe, A. N. Rafferty, and C. D. Manning. Finding contradictions in text. *Proceedings of ACL 2008*, page 1039–1047, 2008.
- [3] S. De Saeger, K. Torisawa, J. Kazama, K. Kuroda, and M. Murata. Large scale relation acquisition using class dependent patterns. In *Proceedings of ICDM 2009*, page 764–769, 2009.
- [4] S.M. Harabagiu, A. Hickl, and V.F. Lacatusu. Negation, contrast and contradiction in text processing. In *Proceedings of AAAI 2006*, pages 755–762, 2006.
- [5] C. Hashimoto, K. Torisawa, S. De Saeger, J.-H. Oh, and J. Kazama. Excitatory or inhibitory: A new semantic orientation extracts contradiction and causality from the web. In *Proceedings of EMNLP 2012*, 2012.
- [6] D. Kawahara, K. Inui, and S. Kurohashi. Identifying contradictory and contrastive relations between statements to outline web information on a given topic. In *Proceedings of COLING 2010*, page 534–542, 2010.
- [7] J. Kazama and K. Torisawa. Inducing gazetteers for named entity recognition by large-scale clustering of dependency relations. *Proceedings of ACL 2008*, page 407–415, 2008.
- [8] J. Kloetzer, S. De Saeger, K. Torisawa, M. Sano, C. Hashimoto, and J. Gotoh. Large-scale acquisition of entailment pattern pairs. In *Information Processing Society of Japan (IPSJ) Kansai-Branch Convention*, 2013.
- [9] K. Murakami, E. Nichols, S. Matsuyoshi, A. Sumida, S. Masuda, K. Inui, and Y. Matumoto. Statement map: assisting information credibility analysis by visualizing arguments. In *Proceedings of the 3rd workshop on Information credibility on the web*, page 43–50. ACM, 2009.