

単語の概念構造と格を考慮したテキスト含意関係認識

阿部建* 栗原正仁 小山聡 佐藤晴彦

(北海道大学 大学院情報科学研究科)[†]

1 はじめに

1.1 研究背景

近年、情報メディアの発達に伴い、個人が取得できる情報が爆発的に増えている。これらの情報はその豊富な量と種類から、様々な解析や知識源として非常に有用であると考えられている。一方で、それらの情報は自然言語で記述されていることも多く、その複雑さ、曖昧さにより効率的な情報の抽出は難しい。

こういった背景から、より深層的な自然言語の意味をとらえるためのタスクとして、テキスト含意関係認識が注目を浴びている。テキストにおける含意関係とは、あるテキストペアについて、前提となるテキスト T1 を見たとき、仮説となるテキスト T2 をおおよそ正しいと推論できるような関係をいう。この技術は様々な情報アクセス技術に適用されることが期待されている。例えば情報検索におけるクエリ拡張や、文書要約における冗長部分の除去などが挙げられる。

1.2 先行研究

日本語におけるテキスト含意関係認識について、NTCIR-9[1]において RITE[2] という評価型ワークショップが初めて実施された。テキストにおける含意関係を示す言語現象には複数の種類があるが、2013 年に実施された評価型ワークショップ RITE-2[3] では、含意を示す言語現象を 22 種類、含意を示さない言語現象を 5 種類に分類している。また、同様に言語現象の関係ラベルを付与する研究として金子ら [4] や小谷ら [5] の研究がある。

Table 1 に、本研究で用いた小谷らによる日本語 TextualEntailment 評価データ [6] に付与されている言語現象の例を示す。本研究ではこのうち、「語彙 (体言)」カテゴリにおける「上位→下位」「下位→上位」「同義語」を扱った。

単語の概念構造を用いた研究として森田らの研究 [7] がある。彼らは日本語 WordNet[8] の階層構造を用いて単語間類似度を算出し、これをもとに閾値を設定して含意を判定した。一方で上位下位関係の方向について考慮されておらず、また単語の構文的役割についても言及していない。

文構造と上位下位関係の方向について着目した研究として柴田ら [9] の研究がある。彼らはテキストを述語項構造として解析し、述語と表層格が一致した項について比べ、仮説に含まれる項全てが前提に含まれる項の上

Table 1

メインカテゴリ	サブカテゴリ
包含	節
	並列
	名詞句
語彙 (体言)	定義的
	同義語
	下位→上位
	上位→下位
語彙 (用言)	名詞句の格関係
	含意
	副詞
	対義語
構文	主語の変換
	複文の変換
	強調構文
推論	結果→原因
	副助詞+一般化
	時間軸・数量

位であったときに含意関係があると判定している。一方で、表層格の情報を比較のすり合わせにしか用いておらず、また単語の概念の階層構造と含意の関係について、一方向しか許していない。

日本語 TextualEntailment 評価データに含まれる含意関係の例を Table 2 に表す。殆どのデータは 1 つ目と 2 つ目のような例が多かった。すなわち、前提が下位、仮定が上位となる場合に含意が成り立つことが多い。他方で、3 つ目から 5 つ目のようなそうではない例外もあった。本研究では先の手法では分類が困難な、3 つ目から 5 つ目の例を分類することを試みる。

Table 2

含意あり	T1	フリッパーはイルカだ	下位語
	T2	フリッパーは哺乳類だ	→上位語
含意なし	T1	彼は女性に会った	上位語
	T2	彼は母親に会った	→下位語
含意あり	T1	一般に小型の哺乳類は冬眠する	上位語
	T2	一般にリスは冬眠する	→下位語
含意あり	T1	文房具を使った	上位語
	T2	定規を使った	→下位語
含意なし	T1	太郎はイルカが好きだ	下位語
	T2	太郎は哺乳類が好きだ	→上位語

2 提案手法

2.1 概要

本研究では含意の方向と上位下位関係の方向の関連を表す特徴をヒューリスティック的に設計し、これを特徴量として機械学習を行った。以降で述べる特徴量は全て構文解析による述語項構造に基づいて、述語及び表層格の一致した項ごとに算出する。述語及び項ごとの特徴量については格単語の平均をとり、学習機に利用する特徴量はこれらのうち最小のものを用いる。

以降の特徴量は全て前提となるテキスト T1 の単語 t_1 、仮定となるテキスト T2 の単語 t_2 が与えられたとき式を表す。また、単語が複数の synset に属している場合、最も値の高くなる組み合わせを採用した。ここで、synset とは日本語 WordNet において単語が属する同義語の集まりであり、この synset を 1 つの概念として、階層構造が構成されている。

含意関係のための特徴値として以下を設計した。

2.2 特徴量

2.2.1 設計コンセプト

日本語では、主語において特に明示しない限りその語が持つ集合全体を指すことが多いと仮定し、この場合その語の性質を表す文章であると考え、Table 2 の 3 つ目の例のように上位語から下位語の方向へ含意が成り立つと考えた。また、ある上位語と下位語が与えられたとき、その上位語から下位語が想起されやすいと、同義的に扱われ、逆に想起され難いと関係性がないと判断され、Table 2 の 4 つ目、5 つ目の例のような含意関係が導かれると考えた。これらの仮説をもとに以下の 4 つの特徴量を設計した。

2.2.2 仮説の単語 t_2 がガ格に属する

ガ格に属する	1
それ以外	0

2.2.3 上位下位関係

t_1 が t_2 の上位・下位関係ともに持つ または 同義語である	3
t_1 が t_2 の下位語である	1
t_1 が t_2 の上位語である	2
上位下位関係を持たない	0

2.2.4 LCS から仮説に対する下位概念の数の逆数

$$\frac{1}{\text{NumOfSyn}(LCS_{t_1,t_2}, \text{Path}(LCS, t_2))}$$

ここで、 LCS_{t_1,t_2} は t_1 と t_2 の共通の synset のうち、最も階層の深さが深い synset を指す。また、 d_x は x の階層の深さを指す。

関数 $\text{Path}(LCS_{t_1,t_2}, t_2)$ は LCS_{t_1,t_2} から t_2 のパス長を示す。関数 $\text{NumOfSyn}(S, \text{path_length})$ はある synset S から path_length だけ下位の位置にある synset の数を示す。

2.2.5 方向性を考慮しない類似度

$$\frac{2 \times d_{LCS_{t_1,t_2}}}{(d_{t_1} + 1) + (d_{t_2} + 1)}$$

3 実験

3.1 実験設定

本研究では単語の概念構造の知識源として日本語 WordNet を用いた。

実験に用いたデータセットは日本語 TextualEntailment 評価データのうち、関係カテゴリが「語彙(体言)」, サブカテゴリが「上位→下位」「下位→上位」「同義語」のものである。ただし、日本語 WordNet に存在しない単語を含むデータは除外するか、英訳して近い単語に置き換えて利用した。また、複数の述語から成り立っているものは除外するか、一方を削除して利用した。データ数は 156 である。

機械学習には SVM-Light, カーネルは RBF, パラメータはデフォルトで行った。

評価については交差検定を leave-one-out 法で行った。

また、比較のためのベースラインとして、1-gram, 2-gram, 3-gram, 4-gram, 形態素重複率の 5 種類の表層情報のみを用いたものと、Table 3 のように t_1 が t_2 の下位語であるという一方向のみを考慮したもので実験を行った。

Fig. 1 F-measure

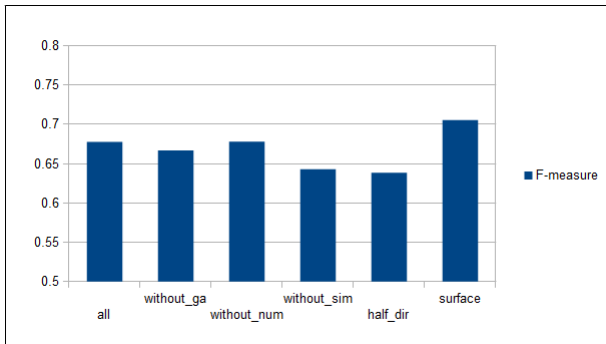


Fig. 3 Precision

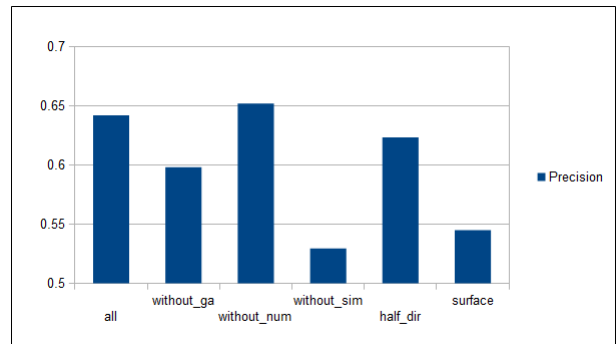


Fig. 2 Accuracy

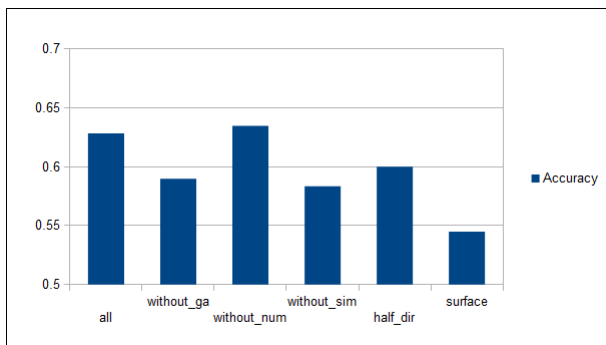
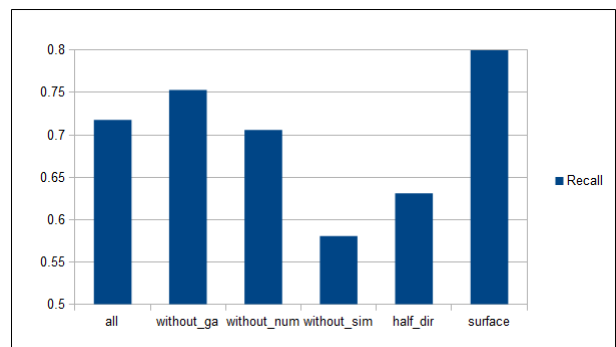


Fig. 4 Recall



3.2 実験結果

実験の結果を Table 4 に示す。また、Accuracy, F-measure, Precision, Recall についてのグラフそれぞれ Fig. 2 から Fig. 4 に示す。実験の結果を Fig. 1 から Fig. 4 に示す。

all は特徴量 2.2.2 から 2.2.5 を用いた結果である。without_ga から without_sim は all からそれぞれ特徴量 2.2.2 から 2.2.5 を除いた結果である。half-dir は all の特徴量 2.2.3 を Table 3 に差し替えた結果である。surface は表層情報のみの特徴量として用いた結果である。

Table 3

t1 が t2 の下位語である	1
それ以外	0

Table 4

	Acc.	F-meas.	Pre.	Rec.
all	0.6282	0.6778	0.6421	0.7176
without_ga	0.5897	0.6667	0.5981	0.7529
without_num	0.6346	0.6780	0.6522	0.7059
without_sim	0.5833	0.6429	0.5294	0.5806
half_dir	0.6000	0.6386	0.6235	0.6310
surface	0.5449	0.7054	0.5449	1.0000

3.3 考察

all と half_dir の結果を見ると、上位下位関係の方向と含意の関係には一方向のみでは不十分であると考えられる。特に、Recall についての向上が顕著にみられ、上位語から下位語の方向において正例となるべきものが上手く分類されたと考えられる。

一方で、all と without_num の結果を見ると、下位概念の数の導入について Recall の値がやや向上しているものの、全体の性能には大きく変化がない。これは下位概念の数の導入の動機として、単語間の想起しやすさを想定していたが、これは人の感覚によるものが大きく、利用したデータセットではその粒度に指標がなかったためであると予想される。

また、all と without_ga の結果を見ると、格の考慮について Precision の向上と Recall の減少がみられた。格の考慮では下位語から上位語の方向で含意が成り立たない例、及び上位語から下位語で含意が成り立つ例の判定を期待していたが、具体的な分類例を見ると期待された判定例は全くなかった。all のみでうまく分類できた例、及び without_ga のみでうまく分類できた例ともに、その殆どが日本語 WordNet 上で上位下位関係を上手く取得できておらず、学習が行えていなかったと考えられる。

all と without_sim の結果を見ると、Precision, Recall ともに向上が見られ、完全一致による単純な文字列比較では含意の判定に不十分であることが考えられる。

surface の結果を見ると、全てのデータを正例に分類す

る結果となり，含意関係の特徴を表しているとは言えない結果となった．これは利用したデータセットの特徴によるものであると考えられる．

4 おわりに

4.1 結論

提案した特徴量は表層情報のみ比べ一定の有効性があると考えられる．一方で，概念構造の利用においては synset の同定が重要であり，性能の向上はその精度に依ると考えらえる．

4.2 今後の課題

本研究で誤判定された例のほとんどが，synset がうまく取得できていなかったものであった．その具体的な理由として，1つは日本語 WordNet にない概念であったことが挙げられる．この解決法として，単純な知識源の拡充か，あるいは応用先に関して特別な知識源を生成することが考えられる．もう1つは synset の同定を誤っていたことが挙げられる．この解決法として，前後の単語との類似を見ること考えられる．また利用したデータセットでは，上位語から下位語の方向へ含意がなりたつ例や，下位語から上位語の方向で含意が成り立たない例が極めて少数であったため，これらを分類することに意味があるかの調査も踏まえ，訓練データ数を増やす必要がある．

参考文献

- [1] NTCIR-9,
<http://research.nii.ac.jp/ntcir/ntcir-9/index.html>, 2012.
- [2] Hideki Shima, Hiroshi Kanayama, Cheng-Wei Lee, Chuan-Jie Lin, Teruko Mitamura, Yusuke Miyao, Shuming Shi and Koichi Takeda. Overview of NTCIR-9 RITE: Recognizing Inference In TExt. Proceedings of NTCIR-9 Workshop Meeting, December 6-9, 2011, Tokyo, Japan.
- [3] Yotaro Watanabe, Yusuke Miyao, Junta Mizuno, Tomohide Shibata, Hiroshi Kanayama, Cheng-Wei Lee, Chuan-Jie Lin, Shuming Shi, Teruko Mitamura, Noriko Kando, Hideki Shima, Kohichi Takeda. Overview of the Recognizing Inference in Text (RITE-2) at NTCIR-10. Proceedings of the 10th NTCIR Conference, June 18-21, 2013, Tokyo, Japan.
- [4] 金子貴美, 宮尾祐介, 戸次大介. 基本文関係に分解した含意関係認識日本語評価データの構築. 言語処理学会 第 19 回年次大会, pp.928-931, 2013.
- [5] 小谷通隆, 柴田知秀, 中田貴之, 黒橋禎夫. 日本語 Textual Entailment のデータ構築と自動獲得した類義表現に基づく推論関係の認識. 言語処理学会 第 14 回年次大会, pp.1140-1143, 2008.
- [6] 日本語 Textual Entailment 評価データ,
<http://nlp.ist.i.kyoto-u.ac.jp/index.php?Textual%20Entailment%20%E8%A9%95%E4%BE%A1%E3%83%87%E3%83%BC%E3%82%BF>
- [7] Hirotaka Morita and Koichi Takeuchi. Construction of a Simple Inference System of Textual Similarity (oka1 RITE2), 2013.
- [8] 日本語 WordNet.
<http://nlpwww.nict.go.jp/wn-ja/>
- [9] Tomohide Shibata, Sadao Kurohashi, Shotaro Kohama, and Akihiro Yamamoto. Predicate-argument Structure based Textual Entailment Recognition System of KYOTO Team for NTCIR-10 RITE-2, 2013.
- [10] rancis Bond, Timothy Baldwin, Richard Fothergill, and Kiyotaka Uchimoto. Japanese semcor: A Sense-tagged Corpus of Japanese. The 5th International Conference of the Global WordNet Association (GWC-2010), pp.9-16, 2010.