

Natural Logic を用いた数量に関する含意関係認識への取り組み

辻有万里[†]小林一郎[‡][†]お茶の水女子大学理学部情報科学科[‡]お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース

1 はじめに

含意関係認識とは、ある文から他の文への推論可能性を自動的に認識することである。含意関係認識は、情報検索、質問応答、テキスト要約など、さまざまな応用分野に必要であり、大きな注目を集めている。このような理由から、これまで数々の手法が提案されてきたが、数量表現を伴う含意関係認識に関する研究はまだ少ない。数量の大小知識を構築し、それに基づいて含意関係を導くという方法が、成澤ら [4] によって提案されているが、この方法で実現するには数量を解釈するための膨大な知識が必要になる。そこで、本研究では、数量を定性的な言語表現に変換できるファジィ理論と自然言語のレベルで一つのテキストからもう一つのテキストを推論するための論理体系である Natural Logic とを組み合わせることにより、数量表現を伴う含意関係認識を実現するための方法に対する基礎的検討を行う。

2 Natural Logic を用いた含意関係認識

Natural Logic とは、自然言語のレベルで一つのテキストからもう一つのテキストを推論するための論理体系である。含意関係認識において、一階述語論理で推論を行う際に自然言語への翻訳の段階で精度が落ちるという問題点を解決するために、MacCartney ら [1] によって提案された。Natural Logic では、単語間において含意関係が保たれるならばその単語間の関係は単調増加であり、否定表現や条件のように、文の含意関係を反転させる単語を単調減少と定義している。このように単語や文法的特徴の単調性を見ることにより、含意関係認識を捉えることができる。日本語文法における Natural Logic の取り組みは、増田ら [2] によって行われている。

An Approach to Recognizing Textual Entailment involving Numerical Expressions using Natural Logic

[†]Yumari TSUJI(g1020523@is.ocha.ac.jp),

[‡]Ichiro KOBAYASHI(koba@is.ocha.ac.jp)

[†]Dept. of Information Sciences, Faculty of Science, Ochanomizu University, 2-1-1 Ohtsuka Bunkyo-ku Tokyo 112-8610

[‡]Advanced Sciences, Graduate School of Humanities and Sciences, Ochanomizu University, 2-1-1 Ohtsuka Bunkyo-ku Tokyo 112-8610

3 数量を捉えた含意関係認識

Androutsopoulos ら [3] による含意関係認識研究のサーベイから、含意関係認識については多くのアプローチがなされているが、数量表現に着目しているものは数が少ないことがわかる。多くの研究では数量表現は他の単語と同様に品詞が割り当てられて処理されているが、含意関係を正しく認識するためには数量の大小判断が必要な場合が多い。数量の大小判断には人間の主観や状況などが大きく関わっているため、含意関係認識の精度を上げるのは困難である。成澤ら [4] は、「も」「しか」「わずか」といった手がかり表現に基づき、たくさん文から数量が表す対象ごとの大小知識を収集している。この研究では収集した知識を元に数量表現を伴う含意関係認識を実現する方法、および知識の必要性について示している。

4 提案手法

Natural Logic に関する先行研究 [1][2] を踏まえ、数量に関する知識の拡充のみに頼らず数量表現を伴う含意関係認識を実現するための手法を提案する。先行研究 [1][2] では数量表現に重点を置いていないため、数量表現に依存する周囲の単語については関係性を反映していない。そこで本手法では、数量表現固有の動詞について数量表現と一緒に考えることで含意関係認識の精度を上げることを目標とする。

文間の含意関係をとる際に必要となる前処理と単語どうしのアライメントは増田ら [2] と同様の手法を用い、Natural Logic 適用部について変更を行う。また、数量表現の曖昧性に対応できるように、ファジィ言語変数を用いる。数量表現が表す対象についてのメンバーシップ関数に基づき、数量表現それぞれにファジィ言語値を割り当てる。さらに、ファジィ演算による類似度を算出する。

先行研究 [1][2] では別々に扱われている数量表現と動詞をセットで一つの文節とし、単調性のタグ付けをし直す。数量表現は単調増加の単語であると考えられるので、動詞の単調性を採用する。ファジィ演算の類似度が高い数量表現に後続する場合の動詞の含意関係を

知識として持っていることを前提として推論を進める。

5 適用例

含意関係の定義は、増田ら [2] のものを採用する。表 1 のように含意関係を定め、文節どうしの含意関係の組み合わせから表 2 に従って含意関係を更新し、推論を進める。表 2 の括弧内の関係は、係り先の文節の単調性が単調減少であるときの含意関係を表している。

表 1: 含意関係の定義

t1 → t2			
t2 → t1	○	×	
	○	=	>
	×	<	?

表 2: 含意関係の組み合わせ

受け側の含意関係	=	<	>	?
係り側の含意関係	=	<<(>)	<(<?)	?(>)
=	=	<	>	?
<	<<(>)	<(<?)	?(>)	?
>	><(>)	?(<?)	>(<?)	?
?	?	?	?	?

ここでは、(i) 「～割」と「～%」などいずれも割合を表す数量表現のように同種のものの場合、(ii) 順位と割合などそれらのみではいかなる判断もできないような異種のものの場合、(iii) 一方が数量表現ではなくファジィ表現になっている場合に分けて考えることにする。各図中の上向き矢印は単調増加、下向き矢印は単調減少を表す。

(i) 同種の数量表現の場合

- t1: チャイルドシートの使用率が 5 割を切った
- t2: チャイルドシートの使用率が 47 % に減った

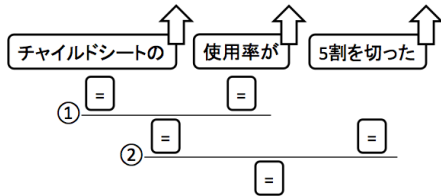


図 1: 同種の数量表現の場合

(ii) 異種の数量表現の場合

- t1: 山梨県はミネラル水の生産シェアが 1 位である
- t2: 山梨県はミネラル水の生産シェアが 50 % を占める

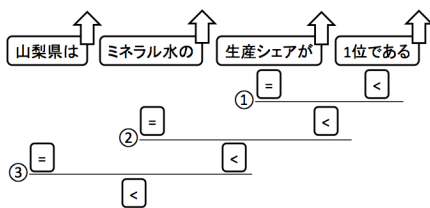


図 2: 異種の数量表現の場合

(iii) 一方が数量表現でない場合

- t1: 友人に少し借金をした
- t2: 友人に 200 円借りた

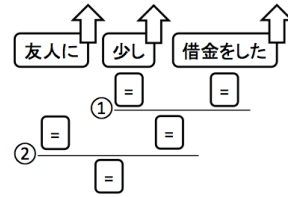


図 3: 一方が数量表現でない場合

5.1 考察

提案手法を適用して推論を行った結果、すべての場合において数量表現を伴う含意関係が導出できた。用いるメンバーシップ関数を工夫することで、人間の感覚に即した推論に近づけていくことが可能だと考える。また、知識として保持している必要がある語彙数も比較的抑えられると考えられる。

6 おわりに

単調性を用いて自然言語のレベルで含意関係の推論を行うことができる Natural Logic を、数量表現を扱えるように拡張した。数量表現ひとつ一つについて、その大小知識を蓄積していくことで含意関係認識を行うのには限界があるため、ファジィ言語変数を用いることで効率よく数量を扱えるようになった。しかし、数量が表す対象ごとにメンバーシップ関数を用意する必要があり、数量表現特有の動詞についての知識は持っていなければならないため、さらに工夫をして効率化をはかることが今後の課題であると考えている。

参考文献

- [1] Bill MacCartney, Christopher D. Manning, Natural Logic for Textual Inference, In ACL-07 Workshop on Textual Entailment and Paraphrasing, 2007.
- [2] 増田涼良, 杉本徹, Natural Logic を用いた日本語テキストの含意関係認識, 第 26 回 人工知能学会全国大会 予稿集, 4K1-OS-2-3, 2012.
- [3] Ion Androutsopoulos, Prodromos Malakasiotis, A Survey of Paraphrasing and Textual Entailment Methods, Journal of Artificial Intelligence Research, Vol.38, pp.135-187, 2010.
- [4] 成澤克麻, 渡邊 陽太郎, 水野淳太, 岡崎直観, 乾健太郎, 数量の大小の自動判定: 「彼は身長が 2m ある」は高いか低いか, 言語処理学会 第 19 回年次大会 発表論文集, pp.354-357, 2013.