

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 (指導教員：小林一郎)

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 [5] 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で、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った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、ある単語をテキスト対の同じ位置に同じように付け加えたとき、含意関係が保たれるならばその単語を単調増加、否定表現や仮定表現のように、含意の方向が反転するならばその単語を単調減少と定義している。このように単語の単調性を考慮することにより、含意関係認識を正しく行うことができる。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。

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。成澤ら [4] は、高精度な含意関係認識システムを構築するためには、数量表現の問題を解決することが必要であるとし、数量表現を伴う文における含意関係認識にどのような課題があるのかを分析している。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 [5] は「も」「しか」「わずか」といった手がかり表現に基づき、たさんの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、収集した知識をもと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言語値を割り当てる。さらに、ファジ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。ファジ演算による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 1 は、例えば「 $t1 \ t2:$ 」すなわち $t1$ から $t2$ が真であると推論でき、かつ「 $t2 \ t1:$ 」すなわち $t2$ から $t1$ が真であると推論できるとき、その含意関係を「 $=$ 」と定義するというを示している。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

$t1 \rightarrow t2$	○	×
$t2 \rightarrow t1$	○	×
○	=	>
×	<	?

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	=	<	>	?
=	=	<	>	?
<	<(>)	<(<?)	?(>)	?
>	>(<)	?(<)	>(<?)	?
?	?	?	?	?

含意関係の組み合わせ方は、表 2 に従って、単調性を考慮しながら、受け側でない文節から係り先の文節へ組み合わせしていく。各テキストの文全体の含意関係を導出したら、それらを同様の方法で組み合わせ、導出された含意関係を 2 つのテキストの最終的な含意関係とする。その際、係り側、受け側、単調性は考慮しない。

ここでは、(i)「 \sim 割」と「 $\sim\%$ 」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。以下、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。

ここでは、それぞれの例について、 t_1 の文全体の含意関係を導出する過程を示す。 t_2 についても同様の手順で導出を進め、2つのテキストの最終的な含意関係を得る。

(i) 同種の数量表現の場合

- t_1 : チャイルドシートの使用率が5割を切った
- t_2 : チャイルドシートの使用率が47%に減った

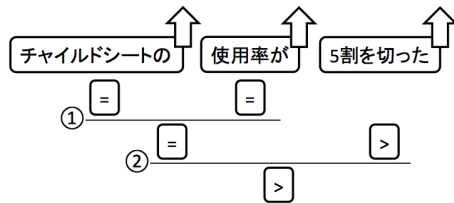


図 1: 同種の数量表現の場合

(ii) 異種の数量表現の場合

- t_1 : 山梨県はミネラル水の生産シェアが1位である
- t_2 : 山梨県はミネラル水の生産シェアが50%を占める

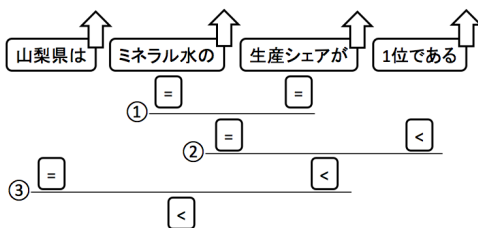


図 2: 異種の数量表現の場合

(iii) 一方が数量表現でない場合

- t_1 : 友人に少し借金をした
- t_2 : 友人に200円借りた

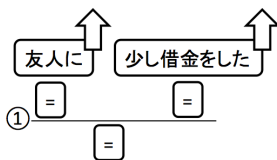


図 3: 一方が数量表現でない場合

5.1 手がかり表現

数量表現は、しばしば手がかり表現を伴う。ここでいう手がかり表現とは、数量表現の前または後について、当該数量表現の大小判断を助ける働きをするものである。表3は、手がかり表現の例を示したものである。例えば「わずか」という手がかり表現は数量表現の前に置かれ、その数量表現が小さいと判断すべきであることを示している。

前述の数量表現と動詞をセットで考える場合と異なり、数量表現を単独で扱う場合に関しても、ファジィ理論を用いることが有効であると考えているが、手がかり表現が付随する場合にはそれだけを考慮して数量表現の大小を判断すればよい。

表 3: 手がかり表現の例

手がかり表現	たった	わずか	しか	だけ	も
数量表現の前/後	前	前	後	後	後
数量表現の大/小	小	小	小	小	大

5.2 考察

提案手法を適用して推論を行った結果、提示した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ることがわかった。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。数量表現一つひとつについて、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いならぬため、さらに工夫をして効率化を図ることが今後の課題であると考えている。メンバーシップ関数については、ファジィ限量子を用いることにより保持すべき関数を減らすことも可能であると考えている。また、特殊な数量表現についても考察する必要がある。例えば、「一つもない」や「一人きり」などのように、1は他の数とは別に扱う必要があり、否定表現に置き換えられることが多い0についても考えていくつもりである。

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, Prodrinos 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, 数量表現を伴う文における含意関係認識の課題分析, 言語処理学会 第 18 回年次大会 発表論文集, pp.1087-1090, 2012.
- [5] 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.