

n-gram を用いた日本語テキスト含意認識の検討

宇高 邦弘, 山本 和英

長岡技術科学大学 電気系

E-mail: {udaka,ykaz}@jnlp.org

1 はじめに

近年テキスト含意認識と呼ばれるタスクが研究され始めている。テキスト含意認識とは、テキストと仮説と呼ばれる二つの言語表現が存在するときテキストの示す意味が仮説の示す意味を含むかどうかを判定するタスクである。以下にテキスト含意認識の例を示す。

テキスト: 坂口安吾は『信長』や『白痴』などの小説を書いた。

仮説: 坂口安吾は小説家である。

含意判定: 含意

テキスト含意認識を用いることで、自然言語処理技術における意味理解という高次の問題を解決することができる。これによって、翻訳精度の指標、質問応答システムでの答えの抽出など、自然言語処理技術の様々な分野に応用することができる。このタスクを実現するにあたって、永田ら [1] は言語表現の事態を理解するための述語項構造解析や言い換えを認識するための語彙知識、ある事態から異なる事態を得るための論理的推論などが必要であると述べている。これらの技術は複雑な処理を必要とするため実現するのが困難である。また、今まで表層的な処理でどこまで含意認識可能であるかについては議論されていない。そこで本稿では、n-gram を単位としてテキストと仮説の一致を見ることでテキスト含意認識を行う。評価セットを用いて実験を行い、本手法でどの程度含意認識が可能か、および本手法でテキスト含意認識を行う上で生じる問題を検証する。

2 関連研究

テキスト含意認識の研究として、日本では小谷らの研究がある。小谷らは国語辞典などから自動獲得した類義表現を用いて推論パターンを構築し、含意認識を行った [2]。また、言語表現を述語項構造に正規化し、述語項構造を単位としてテキストと仮説間の一致を見ることで含意認識を行うシステムを構築している [3]。加えて述語や格要素の上位下位関係の認識や、否定表現および仮定法過去を用いた表現に対する例外処理も行っている。海外では Andrew and Jeremy の研究がある。Andrew and Jeremy はいくつかの意味関係に基づいて含意文生成規則を作成し、その規則によってテキストから複数の含意文を生成することで含意認識を行っている [4]。しかし、これらの処理は格フレームの獲得や語句の同義関係や上位下位関係および反義関係の認識、生成規則の作成という複雑な処理を必要とする。本研究は n-gram を用いることで複雑な処理を使用しないテキスト含意認識を目指す。本手法はテキストと仮説を決められた単位に分解し両文の一致をみることで含意認識を行うため、格関係の同定などの処理を必要としない。また、n-gram は要素を決められた要素数でつなぎ合わせることで作成可能であるため処理が容易である。

3 n-gram を用いたテキスト含意認識

n-gram の要素として形態素、形態素から助詞、助動詞を除いたもの (以下、自立語)、文節の 3 つを用いた。形態

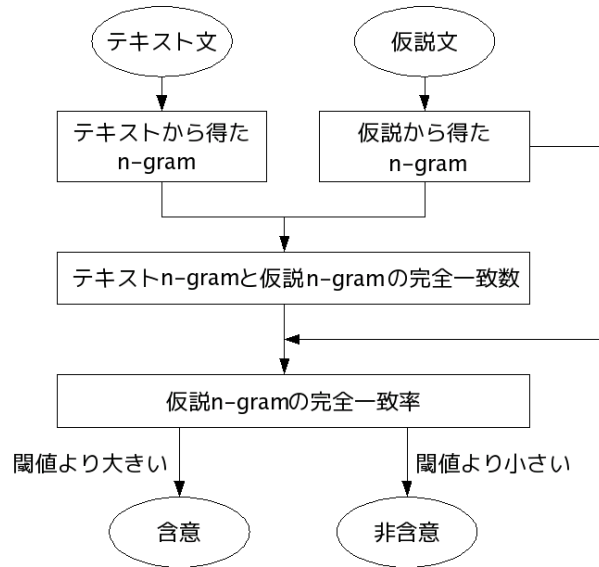


図 1: 本手法における含意認識の流れ

素および自立語 n-gram の作成時には形態素解析器「茶筌⁽¹⁾」を使用し、文節 n-gram 作成時には構文解析器「南瓜⁽²⁾」を使用した。また、n-gram の要素数は 1 から 3 までを用いた。

3.1 テキスト含意認識の流れ

テキスト含意認識は次の流れで行った。

1. テキストと仮説両文から同じ要素、要素数で n-gram を作成
2. 作成したテキストと仮説の n-gram を比較し、完全一致する n-gram 数を計算
3. 仮説の n-gram 総数を h 、2 で得た完全一致する n-gram 数を a 、含意するかしないか判定するための閾値を t としたとき、含意判定は式 (1) で表される。このとき、閾値 t は 10~100 の値を取る。また、この式において $\frac{a}{h} * 100$ を仮説 n-gram の完全一致率とする。

$$\begin{cases} \text{含意} = \frac{a}{h} * 100 \geq t \\ \text{非含意} = \frac{a}{h} * 100 \leq t \end{cases} \quad (1)$$

以下に、本手法での含意認識例を示す。

n-gram を用いた含意認識例

要素: 形態素 要素数: 2 閾値 t : 50

テキスト: 太郎に本を借りたが無くした

n-gram: 太郎に に本 本を を借り 借りた たが
が無く 無くし した

仮説:太郎に本を借りた
n-gram:太郎に に本 本を を借り 借りた

完全一致 n-gram 数 a:5
仮説 n-gram 数 h:5
含意判定:含意

この例では、仮説 n-gram 全てがテキストと完全一致し、仮説 n-gram の完全一致率が閾値より高くなるので含意していると判断される。

4 評価実験

4.1 テストセット

本手法について検証するためにテストセットを作成した。テストセットはテキスト 1 文と仮説 1 文のペア 104 個で構成される。これらのテキストは日本経済新聞 1990 年～1999 年から無作為に抽出した 1 文を使用し、仮説は PASCAL が公開している含意認識テストセット⁽³⁾を参考に、人手で作成した。多くの仮説は、ペアとなるテキストで使用されている語句を 5 割以上含んで構成されている。また、テキストと仮説が同文であるペアは存在しない。テキストは平均して 30 形態素で構成され、仮説は平均して 11 形態素で構成されている。テキストと仮説の各ペアにはあらかじめ人手で含意か非含意かのタグが付与されており、これを正解データとした。正解データの例を以下に示す。

正解データの例

<t>愛知県警春日井署はひき逃げ事件として、逃げた車の行方を追っている。</t>
<h>新潟県警はひき逃げ犯を追っている </h>
<RTE>x </RTE>

含意ペアと非含意ペアは 52 個ずつ用意した。また含意認識時、閾値は 10～100 まで 10 刻みに設定した。システムが出力した解答と正解データを比較し、各要素、要素数、閾値において正しく判定できているかを検証した。

4.2 実験結果

含意ペアのみでの実験結果を図 1～図 3 に、非含意ペアのみでの実験結果を図 4～図 6 に示す。図 1～図 6 において、精度とは式 (2) で表される。

$$\text{精度} = \frac{r}{p} \quad (2)$$

式 (2) において、r は本システムにおいて含意判定した結果が正解データと一致したペア数を示し、p は全ペア数である。また、表 1 に、各要素および要素数において、閾値を変化させても不正解となるペア数を示す。

5 考察

5.1 本稿のシステムに関して

図 1～図 6 から、要素、要素数、閾値の変化において以下のような傾向を得た。

- 要素
形態素:含意と判定する傾向にある
自立語:含意と判定する傾向にある
文節:非含意と判定する傾向が強い
- 要素数
小さいほど含意と判定する傾向にある
大きいほど非含意と判定する傾向にある

・ 閾値

低いほど含意と判定する傾向にある
高いほど非含意と判定する傾向にある

5.2 各要素における含意認識の傾向について

表 2 において、形態素および自立語 n-gram は不正解となるペア数に差はあまり見られないが、文節 n-gram は他の要素よりも不正解となるペア数の差が大きい。特に 1-gram では、形態素および自立語 n-gram は非含意となるペアでのみ不正解ペアが存在するが、文節 n-gram では含意となるペアでの不正解が多い。これは各要素における含意認識の傾向が理由である。文節 n-gram は自立語と助詞、助動詞をまとめて 1 つの要素とするため、テキストと仮説で同じ自立語と助詞、助動詞が連続しない場合完全一致しない。そのため、非含意と判断する傾向がもつとも強い。

形態素 n-gram も文節 n-gram と同じく自立語に加えて助詞、助動詞まで一致を見るが、自立語と助詞、助動詞を分けて一致を見るため、文節 n-gram よりも含意と判定する傾向にある。特に形態素 1-gram はテキストと仮説で同じ自立語および助詞、助動詞が連続する必要がなくなるため、含意と判定する傾向が強い。

自立語を要素とすると助詞、助動詞まで完全一致する必要はなくなるため、含意と判定する傾向にある。閾値および要素数の影響で文節および形態素 n-gram は非含意と判定するペアでも、自立語 n-gram では含意と判定できるペアが存在する。以下に、文節 n-gram のみが不正解となる例、形態素および自立語 n-gram が不正解となる例を示す。加えて、形態素および文節 n-gram では非含意と判定されることがあるが、自立語 n-gram では含意と判定される例を示す。

文節 n-gram のみが不正解となる例

テキスト:アンソニー女史は大統領と同郷のアーカンソー州ホープ出身。
仮説:大統領はアーカンソー州ホープの出身である。

形態素および自立語 n-gram が不正解となる例

テキスト:軽トラックは約三十分後に現場から約一キロ離れた同市今の庄二の路上で見付かった。
仮説:軽トラックは約十分後に見付かった。

自立語 n-gram で含意と判定される例

テキスト:一方、欧州では豚コレラや狂牛病が発生。
仮説:欧州で豚コレラが発生した。

5.3 各要素数の含意認識傾向について

表 1 において、形態素および自立語 1-gram において不正解となるペアは、全て非含意ペアであることが分かる。逆に、形態素および自立語 2-gram、3-gram において不正解となるペアは、多くが含意ペアであることが分かる。1-gram での一致は要素が連続する必要が無いいため、1-gram は含意と判定する傾向にある。そのため、テキストと仮説で意味が異なるが構成する要素が同じである場合には含意認識が困難となる。2-gram および 3-gram での一致は要素の出現する順番がテキストと仮説で同じでなければならぬため、2-gram および 3-gram は非含意と判定しやすい傾向にある。そのため、テキストと仮説の両文を構成する要素が同じでも、出現順序が異なる場合には含意認識が困難となる。次に示す例は 1-gram で含意認識が困難である例と、2-gram で含意認識が困難である例である。

表 1:全閾値で不正解となるペア数

要素	1-gram		2-gram		3-gram	
	含意	非含意	含意	非含意	含意	非含意
形態素	0 / 52	24 / 52	10 / 52	0 / 52	18 / 52	0 / 52
自立語	0 / 52	19 / 52	13 / 52	2 / 52	20 / 52	2 / 52
文節	20 / 52	1 / 52	37 / 52	0 / 52	42 / 52	0 / 52

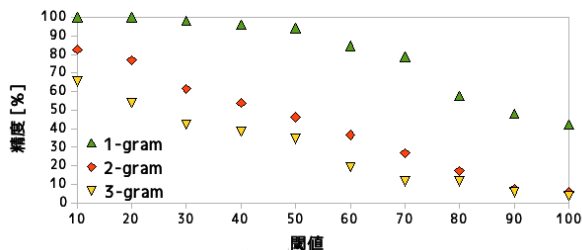


図 1:形態素を要素としたときの含意ペア実験結果

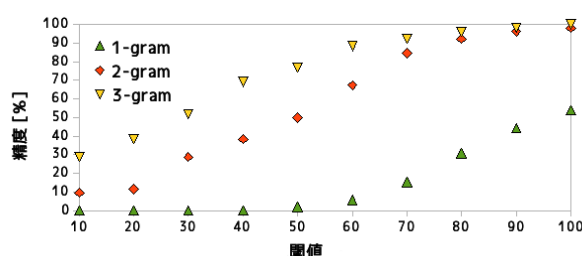


図 4:形態素を要素としたときの非含意ペア実験結果

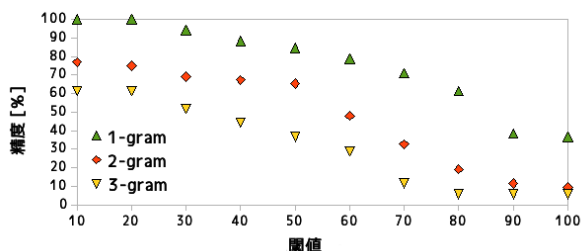


図 2:自立語を要素としたときの含意ペア実験結果

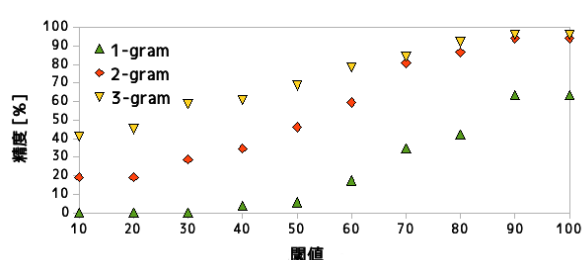


図 5:自立語を要素としたときの非含意ペア精度結果

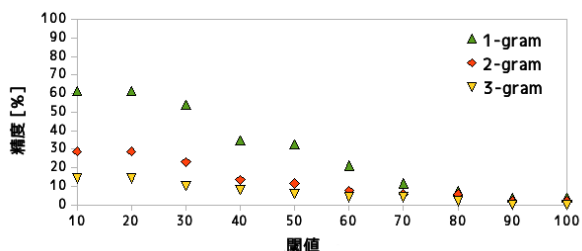


図 3:文節を要素としたときの含意ペア実験結果

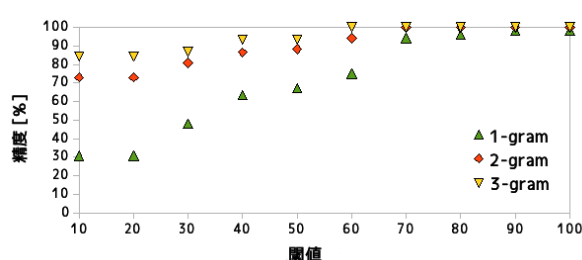


図 6:文節を要素としたときの非含意ペア精度結果

1-gram で含意認識が困難である例

テキスト:同署の調べでは、柴崎容疑者は同日午前七時十分ごろ、肩を骨折して入院していた柳本章さんと口論になり、持っていた包丁で柳さんのわき腹などを刺し、二カ月の傷を負わせた。
 仮説:柴崎容疑者は肩を骨折して入院していた。

2-gram で含意認識が困難である例

テキスト:愛知県警豊田署は二人を刺した同市西岡町長土井一三、運転手、安藤昇容疑者を殺人などの疑いで逮捕した。
 仮説:運転手が殺人事件を起こした可能性がある。

また、各要素における要素数ごとの不正解ペアで一致するものを調べると、次のような傾向が得られた。

形態素および自立語 n-gram

- 1-gram で不正解となるペアと 2-gram で不正解となるペアは一致するものが少ない
- 2-gram で不正解となるペアは 3-gram でも不正解であることが多い

文節 n-gram

- 1-gram で不正解となるペアの多くは 2-gram でも正解しない
- 2-gram で不正解となるペアの多くは 3-gram でも正解しない

表 1 において、文節 n-gram で不正解となるペア数が要素数が大きくなるにつれて増えるのは、文節を要素としたときには非含意と判定する傾向が強いことに理由があると考

える。文節を要素としたときには、1-gram でも多くのペアを非含意と判定する。要素が大きくなるにつれ非含意となる傾向がさらに強くなるため、多くのペアを非含意と判定する。従って、正解が含意であるペアのほとんどを正しく判定できない。

5.4 精度に関して

本稿のシステムにおいて、含意認識精度は要素、要素数、閾値の傾向に依存する。特に閾値の傾向によって大きく依存すると考える。これは式(1)においてテキストと仮説の含意関係を閾値によって判定しているためである。閾値が低ければ、多くのペアを含意と判定するため、含意ペアでの精度はあがるが非含意ペアでの精度は下がる。逆に、閾値が高ければ多くのペアを非含意と判定するため、非含意ペアでの精度はあがるが含意ペアでの精度は下がる。この傾向は、要素および要素数が変化しても変わらないため、精度への影響が大きいと考える。

要素、要素数、閾値ごとの含意認識結果をより正しく含意認識するための判断材料にすることで、本システムはより正しく含意認識可能になると考える。例えば、要素や要素数、閾値それぞれの含意認識傾向をもとに、非含意と判断する傾向にある状況で含意と判断された場合はその結果を重要とし、逆に非含意と判断する傾向にある状況で非含意と判断しても重要としないなどとする。このように含意認識結果それぞれについて重要度を計算し、含意と非含意のどちらが高い重要度を示すかによって含意認識を行うことで、より正確な含意認識が可能であると考えられる。

また、語彙知識の収集も精度向上に効果的であると考えられる。本システムはテキストと仮説が同じ要素で構成されているか否かで含意認識を行っている。そのため、テキストと仮説が同じ意味でも、異なる表現で構成されていると含意と判断できない。語彙知識を用いて、テキストと仮説を構成する語句や表現が換言関係にあるかどうかを認識する。換言関係にあるなら同じ表現に換言していく動作を前処理として行うことで、本稿のシステムはより正しく含意関係を認識できる。加えて、対義語や否定表現を語彙知識を用いて認識することで、さらに精度が向上すると考える。テキストを構成する言語表現が仮説を構成する言語表現と対義関係にあるときには非含意と判定し、否定表現がある場合には含意認識結果を反転することで含意認識を正しく行える。以下に語彙知識を必要とする、本システムでは正しく含意認識できない場合があった例を示す。

語彙知識の必要な含意認識の例

・換言

テキスト: 芭蕉布作りは労多くて骨の折れる仕事です。
仮説: 芭蕉布は作るのが難しい。

・対義語

テキスト: 日本の橋本外務報道官は身代金報道について、「根も葉もないこと」と全面否定している。
仮説: 報道官は身代金報道を肯定した。

・否定表現

テキスト: 岡下さんに大きな怪我はなかった。
仮説: 岡下さんは大きな怪我をした。

加えて、推論システムの構築も、含意認識精度向上につながるかと考える。テキストと仮説で使用される語句が同じであっても、語句の出現順序が異なる場合は不正解となることがあった。以下に示す例は出現順序の変化により不正解となった例であるが、同時に論理的推論を必要とする含

意認識である。

推論

テキスト: 都心の大手ホテルの正月宿泊プランが盛況だ。
仮説: 正月の都心の大手ホテルは盛況である。

この例では「大手ホテルの宿泊プランが盛況である」という事実から、「大手ホテルが盛況である」という論理的推論を必要とする。推論を行うシステムを使用して、テキストを換言し仮説との一致を見ることで含意認識が可能になると考える。

6 おわりに

本稿では簡単な処理で含意認識を行うために、n-gram を単位としてテキストと仮説の一致を見た。また、作成したテストセットを用いて含意認識精度および認識失敗の原因の検証を行った。その結果、本稿のシステムでの含意認識精度は閾値に大きく依存した。また、要素や要素数を変化させたときの含意認識傾向を確認することが出来た。

今後、本システムの結果から得た傾向をもとに、要素や要素数、閾値の変化による含意認識結果の変化を判断材料として含意認識を行うシステムの考案や、文脈、意味まで考慮した語彙知識を収集することで、精度の高い日本語テキスト含意認識方法を研究していく予定である。

使用した言語資源及びツール

- (1) 形態素解析器「茶釜」, Ver.2.3.3, 奈良先端科学技術大学院大学 松本研究室,
<http://chasen.naist.jp/hiki/ChaSen/>
- (2) 構文解析器「南瓜」, Ver.0.52, 奈良先端科学技術大学院大学 松本研究室,
<http://chasen.org/~taku/software/cabocha/>
- (3) Pattern Analysis, Statistical Modelling and Computational Learning ,
<http://pascallin.ecs.soton.ac.uk/>

参考文献

- [1] 永田 昌明, 藤田 早苗, 平 博順. 汎用的な意味解析技術への挑戦. NTT 技術ジャーナル, pp.24-27, 2008.6.
- [2] 小谷 通隆, 柴田 和秀, 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 発表論文集, pp.1140-1143, 2008.
- [3] 小谷 通隆, 柴田 和秀, 黒橋 禎夫. 言い換え表現の述語項構造への正規化とテキスト含意認識での利用. 言語処理学会 第 15 回年次大会 発表論文集, pp.260-263, 2009.
- [4] Andrew Hickl and Jeremy Bensley. A Discourse Commitment-Based Framework for Recognizing Textual Entailment. In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pp.171-176, 2007.