

含意認識における主題に着目した文の比較手法の検討

森田 寛隆 竹内 孔一

岡山大学大学院 自然科学研究科

{morita, koichi}@cl.cs.okayama-u.ac.jp

1 はじめに

含意認識とはテキスト (T) と仮説 (H) と呼ばれる文の対が与えられたときに、仮説の持つ意味内容をテキストが含むか否か、含意の有無を判別するタスクである。日本では NTCIR にて RITE2[1] という含意認識のワークショップが実施されている。

含意認識のデータとしては

T: 伊坂幸太郎は直木賞候補になった 2003 年の『重力ピエロ』で一般読者に広く認知されるようになった。

H: 『重力ピエロ』は伊坂幸太郎による小説で直木賞候補作品だった。

といった文の対とその含意の有無が正解ラベルとして与えられ、この例であれば、含意あり、が正解ラベルとして振られている。

含意認識のように、文の比較を行う際には文の要素、文節や形態素の比較を行う必要があり、とりわけ文の話題の中心となる主題についての記述が 2 文で同様であるかどうかは、文全体が同様の内容を表しているかどうかにか大きな影響があると思われる。文対で主題が一致していれば、同じ対象に関する 2 文として扱うことができ、主題が揃っていない場合よりも容易に 2 文の比較、含意認識が行えると期待できる。

そこで、本研究では、日本語の含意認識タスクを行う際に、仮説の主題にテキストの主題を揃えた構造を考えることで、同じ主題に対する文として 2 文を扱い、比較することで含意認識を行う手法を提案する。主題を揃える操作は文の比較の前処理として行うことができ、各文節や形態素の比較に様々な手法や外部知識を用いることができる。本稿では単純なチャンクの原形の比較のみで含意認識を行い、その精度評価や主題を揃えられなかった例について報告する。

2 関連研究

本研究に関連したものに、意味構造とその変形を行っているものとして、高木ら [2] の研究がある。高

木らは「あの車は赤い」や「赤い色をした車」といった、同じ意味を持っているが表現が異なっている文を同様に扱えるよう意味構造を構築している。本研究では主題が異なる 2 文について、片方の主題にもう一方の主題を揃える操作を行ってその主題に関する文節をぶら下げた構造を作るが、高木らの手法では主題が異なる場合ではなく、同じ主題に関する文を扱う際にどのように構造をつくるべきかを考えている点異なる。そのため、本手法と高木らの手法を組み合わせることも可能であると思われる。

3 提案手法

我々は次のような手順で仮説とテキストの含意認識を行うシステムを構築した。まず、仮説とテキストそれぞれに対し () に囲まれた文字列を削除するなどの前処理を行う。次に、前処理を行った文対に対し係り受け解析器 CaboCha¹ で係り受け解析を行い、機能語が係助詞の「は」であるチャンクを主題として扱う。「は」がない場合は格助詞「が」が機能語であるチャンクを主題として扱う。

そして、仮説の主題と編集距離の近いチャンクをテキストから探し、それを新たなテキストの主題として扱う主題合わせを行う。この際、テキストのもともとの主題と新たな主題について、それぞれにかかっていた修飾語の関係を維持する。新たな主題の機能語を「は」に書き換え、もともとの主題は適切な機能語に置き換える。多くの場合は主語を表す「が」に書き換える。もともとのテキストの主題が仮説の主題と類似していればこの主題合わせは行わない。主題合わせの様子を図 1 に示す。図 1 に表した「(小説だ)」という文節は、実際は挿入されないが、意味が通るものとして表すために書き加えている。

¹<http://code.google.com/p/cabocha/>

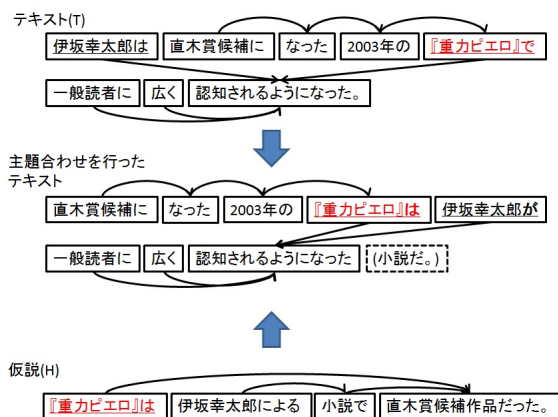


図 1: 主題合わせの例

仮説と、主題合わせを行ったテキストでそれぞれ主題の関わるチャンクを主題にぶら下げた構造を作る。編集距離の近い主題をもつ構造同士が持つチャンクを比較し、一致率を出す。最後に、その一致率が設定した閾値よりも高ければ含意ありとし、低ければ含意なしとする。また、編集距離の近い主題がなかった場合も含意なしとする。構造の比較の様子を図 2 に示す。

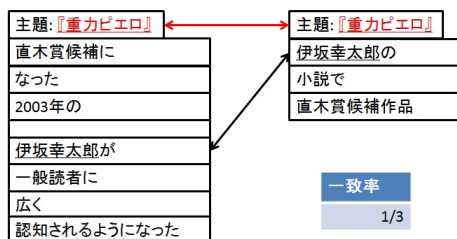


図 2: 構造比較の例

4 評価実験

RITE2 BC サブタスク訓練用データを用いて評価実験を行った。611 文対あり、うち含意関係ありは 271 文対、含意関係なしは 340 文対であった。

4.1 ベースラインシステム

比較のために、CaboCha で係り受け解析を行って得られたチャンクから助詞などを取り去って、その表層を文対で比較するベースラインシステムを構築した。テキストと仮説のチャンクの表層一致の割合が閾値を上回れば含意あり、そうでなければ含意なしと判断する。

4.2 実験結果

表 1 と表 2 に実験結果を示す。表 1 はデータ全体に対する精度と F 値 (含意関係ごとの F 値の平均)、表 2 は含意関係ごとの適合率、再現率、F 値を示している。「主題一致のみ」で表しているのは、提案手法で主題を揃えることが出来なかった文対を除いた場合の結果である。閾値はいずれの手法でもデータ全体の F 値が最も良かった値を採用している。

表 1: データ全体の評価

手法	精度	F 値
提案手法	72.01(440/611)	66.70
主題一致のみ	70.34(230/327)	49.82
ベースライン	71.36(436/611)	70.57

表 2: 含意関係ごとの評価

手法	含意	適合率	再現率	F 値
提案手法	あり	77.17	40.83	53.41
	なし	70.66	92.18	80.00
主題一致のみ	あり	77.17	40.83	53.41
	なし	66.00	35.58	46.23
ベースライン	あり	61.99	70.00	65.75
	なし	78.82	72.24	75.39

提案手法で主題が揃えられたものは 611 文対中 327 文対であり、およそ半分であった。その結果、含意ありの文対に対する適合率はベースラインを上回ったが、再現率では大きく下回った。

5 考察

主題が揃えられなかったものが半分程度あったが、ほとんどの場合は名詞句の主要語をとれなかったことが理由と思われる、主題がないために揃えられなかった例はまれであった。

5.1 仮説に主題がなかった例

主題が仮説になかった例は 5 例のみであり、次のような例があった。

T: 型枠とは、液体状材料を固化させる際に、所定の形状になるように誘導する部材、枠組みのことで、コンクリートや発泡スチロールなどの成形に用いられる。

H: コンクリートを打設する際に型枠を使用する。

この仮説は、場合 + 動作の組み合わせで、行動規範や習慣のようなものを表していると思われる。このような例であれば、動作を表す部分にヲ格がある場合は、そのヲ格 (この例では「型枠」) を主題として扱うとよいように思われる。このような習慣を表す例以外には、過去の経緯を表す例が 1 例だけあった。

5.2 誤った文節を主題とした例

名詞句に主題となるべき要素が含まれていたが、取るべきでない体言を主題として取ってしまった例があった。

T: 看板メニューである焼き餃子はニラよりにんにくを利かせ、たっぷりの肉とキャベツなどを細かく刻み混ぜこんだ餡は、食べた時に肉汁が口の中に広がる。

H: 焼き餃子の具としては、ニンニクの役割を果たすものにニラがあるが、必ずしも入れず、白菜と豚肉のみなどというものも多い。

この文対の共通の主題として選ばれるべきは「焼き餃子」であったが、仮説の主題として「具」を選んできてしまい、主題が揃えられなかった。「具」は身体部分のようなものであり、これ単体では主題となりえず、「焼き餃子」を主題として、その部分属性のように扱うべきである。他にも「上」や「先」といった、位置関係を表す語も主題として扱うべきでないものとして考えられるが、そのような位置関係を表す語を主題として誤って取ってしまった例はなかった。対処法として、このような主題となりえない名詞を辞書に登録する方法が考えられる。

主題の部分文字列は一致しているが、同じ主題として扱えなかった例が多かった。

T: 箱詰めされた夕張メロンおよび夕張キングメロンは、北海道夕張市を生産地とするメロンである。

H: 夕張市は夕張メロンの産地として知られる。

この例では「夕張市」が共通の主題であるが、仮説の「夕張市」とテキストの「北海道夕張市」とチャンクの編集距離でマッチングを行っており、編集距離が大きく、主題のマッチングが取れなかった。こういった挙動が望ましい場合もあり、

T: タブは 1962 年に登場したコカコーラ社初のノンカロリーのコーラ風味飲料である。

H: タブクリアは、かつてコカ・コーラから発売されていたノンカロリーコーラのブランドのひとつだ。

この例では「タブ」と「タブクリア」は異なる固有名詞として扱わなければならないが、部分が一致している

からといって同じ主題としてはならない。商品名などの固有表現かどうか判定したり、形態素単位で、その形態素が主要語となりえるかどうかを考慮したりして主題を扱う必要があると思われる。

主題として時間表現を取ってしまったものもあった。

T: 忠岡町は、大阪府泉北地域に位置する町で、日本一面積の小さい町でもある。

H: 2009 年 10 月 1 日以降は忠岡町が日本一面積の小さい町となった。

この例では機能語が「は」である「1 日以降」を主題として扱っており、「忠岡町」を共通の主題として扱えていなかった。他にも「～のときは」といったチャンクを主題としてとってしまう例もあった。時間表現についても、主題となりえないものとして扱うべきだと思われる。

6 おわりに

本研究では文の主題を揃えた意味構造を比較することで含意認識を行う手法を提案した。単純なチャンクの表層比較による含意認識実験を行い、主題を揃わずにチャンクを比較する場合に比べ、含意のある文対に対する適合率が向上した。しかしながら、主題を揃えることができた文対はデータセット 611 文対のうちの半分程度の 327 対であり、再現率は大幅に低下した。しかしながら、主題とはなりえない名詞を辞書として持つなど、主題を揃えるための外部知識を追加することで性能の改善が期待できる。本手法は主題を揃えて比較する操作しか行っていないため、外部知識や手法を組み合わせることが容易であり、今後は外部知識や文の要素を比較する別手法を加えて評価実験を行う予定である。

参考文献

[1] Yotaro Watanabe, Yusuke Miyao, Junta Mizuno, Tomohide Shibata, Hiroshi Kanayama, C.-W. Lee, C.-J. Lin, Shuming Shi, Teruko Mitamura, Noriko Kando, Hideki Shima, and Kohichi Takeda. Overview of the Recognizing Inference in Text (RITE-2) at NTCIR-10. In *Proceedings of the 10th NTCIR Conference*, 2013.

[2] 高木 朗, 伊東 幸宏. 自然言語の処理. 丸善, 1987.