述語項構造と機能表現の比較と機械学習を併用した文章の含意関係認識手法

伊藤 大紀[†1]　田中 正浩[†1]

駒田 康孝[†2]　鈴木 大地[†2]　山名 早人[†3†4]

†1 早稲田大学基幹理工学部 〒169-8555 東京都新宿区大久保 3-4-1

†2 早稲田大学大学院基幹理工学研究科〒169-8555 東京都新宿区大久保 3-4-1
†3 早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1
†4 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: {ito_d, tanaka_m, ykoma, dsuzuki, yanama}@yama.info.waseda.ac.jp

あらまし 自然言語処理の研究分野の一つである含意関係認識は，自然言語処理に広く共通する基礎的課題である．この課題の精度向上は情報検索や評価分析など，幅広い研究分野において有用である．含意関係認識を行う方法としては，文章間の類似度を特徴量として抽出し，機械学習を用いて判定器を構築する手法が一般的である．しかし機械学習を用いた判定器の構築では，言い換えや比喩などが多く存在し，人工言語に比べ無秩序である自然言語の意味を解析する上で精度が良くないという問題があった．そこで本研究では，機械学習に頼らない分類方法と，機械学習を併用することにより含意関係認識の精度向上を目指す．具体的には，述語項構造の比較による分類方法と，含意関係がない文章対検出による分類方法を加える．評価実験として，NTCIR が提供するデータセットを用いて，500 件の文章対に対して文章対間に含意関係があるか否かの 2 種類に分類する実験を行った．実験の結果，既存の機械学習のみによる分類手法の精度が 58.0%であるのに対し，提案手法では 60.8%であった．含意関係認識問題において，特定の文章に対して述語項構造の比較による分類方法，文章中の機能表現を考慮した分類及び含意関係がない文章対検出による分類方法を適応させることにより，精度が向上することを確認できた．

キーワード 自然言語処理，含意関係認識， NTCIR， RITE

1. はじめに

自然言語処理の研究分野において含意関係認識は近年注目を集めている．含意関係認識とは 2 つの文章からなる文章対のうち，一方の文章からもう一方の文章が推論できるか否かを判定する問題である．含意関係認識の技術が発達すると，異なる文章間の同義性や含意関係，矛盾を正しく認識することができるようになるため，情報検索，質疑応答，文章要約などの幅広い研究分野に活用することができる．

含意関係認識は，欧州連合が資金を提供している研究者ネットワークである Pascal[1]において 2005 年から評価タスクが始められた．2008 年からはアメリカ国立標準技術研究所(National Institute of Standards and Technology)が主催する Text Analysis Conference[2] (以下，TAC)が Pascal を取り込み，2012 年時点で 7 回評価タスクが行われている．日本においては，国立情報学研究所が主催する NTCIR[3]が評価タスクである Recognizing inference in TExt[4] (以下，RITE)を 2011 年度から行なっている．

プログラミング言語などの人工言語と比べて，自然言語には言い換えや比喩，一般常識などによる説明の省略などが多く存在し，比較的に無秩序である．日本語文章の含意関係認識においては，文章間の類似度などを特徴量として抽出し，Support Vector Machine(以下，SVM)による学習を行なって判定機を構築する手法が一般的であるが，比較的に無秩序な自然言語に対しては含意関係を判断する有効な特徴を発見することが困難なため，効率的な手法とは言えない．NTCIR で行われた 2011 年度の RITE では，含意関係があるか否かの分類に対する最高精度が 58.0%と余り高くない．そこで我々は，特定の条件下で含意関係認識が容易な文章対と含意関係認識が難しい文章対とを機械的に分け，それぞれについて別の分類基準を設け，含意関係を判断する手法を提案する．つまり，ある指標によって含意関係認識が容易と考えられる文章対が入力として与えられた場合には，その指標に基づいて判定を行い，含意関係認識が難しいと考えられる文章対に対しては機械学習によって構築した判定機を用いて判定するというものである．具体的な分類手法としては，述語項構造の比較による分類方法，文章中の機能表現や固有名詞を考慮した分類方法及び機械学習による分類を併用させることにより精度向上を目指す．

本稿では以下の構成を取る．まず 2 節では NTCIR の RITE について紹介する．3 節で関連研究についてまと

め，4 節で提案手法について述べる．5 節で評価実験と評価を行う．最後に 6 節でまとめを述べる．

2. NTCIR の RITE について

本研究では，ワークショップ型共同研究である NTCIR の中で文章間の含意関係認識を目的とした RITE タスクの課題を行っており，データも NTCIR から提供されたものを使用している．本節では参加したワークショップである NTCIR と，本研究で行う RITE のサブタスクの 1 つであるバイナリクラスサブタスク(以下 BC サブタスク)について説明する．

2.1. NTCIR と RITE の概要

NTCIR とは，情報検索と，文章要約・情報抽出などの文章処理技術に対する研究の更なる発展を図るワークショップ型共同研究である．NTCIR の目的は大きく分けて次の 3 つであり，テストコレクションと実験結果を評価するための共通手順が用意され，広く提供される点が特徴である．

- 大規模かつ再利用可能なテストコレクションの構築
- システムの比較，研究上のアイデアの交換等に寄与する研究者フォーラムの展開
- 情報アクセス技術の評価手法，評価指標に関する研究の推進

NTCIR には質問応答や文章要約などの様々なタスクがある．本研究を行うにあたり我々が参加した RITE は，文章間の含意，換言，矛盾の認識を目的としたタスクである．

2.2. BC サブタスクの概要

RITE タスクは 4 つのサブタスクからなる．本研究で取り組んだサブタスクである BC サブタスクの説明と評価方法について説明する．

BC サブタスクは，*t1*，*t2* からなる文章対を入力して，*t1* から仮説 *t2* が真だと推論できるか否か，すなわち *t1* が t2 を含意するか否かを判別する 2 値分類問題である．テキスト対の例を図 1 に示す．図 1 の例だと文章 *t1* の内容から文章 *t2* の内容が正しいと推論できるので，*t1* が *t2* を含意するといえる．文章対を入力したときに，*t1* が *t2* を含意する場合は Yes を，含意しない場合は No を出力するようなシステムを作成する．

RITE タスクの評価方法は次の手順で表される．

① 学習用データとして *t1*，*t2* からなる文章対とそれに対する正しい出力結果が XML ファイルで与えられる．
② 参加チームは学習用データや NTCIR が提供するツールなどを用いてシステムを開発する．
③ 評価用データが学習用データと同じ XML ファイルで与えられる．ただし評価用データでは出力結果は記載されていない．
④ 開発したシステムを用いて評価用データから各文章対に対しての出力結果を記載したファイルを提出する．
⑤ 評価用データの正解に対する出力結果の正答率が評価の対象となる．

以下，⑤で述べた評価用データの正解に対する出力結果の正解率を精度と呼ぶ．与えられる XML ファイルの例として，BC サブタスクの学習用データを図 2 に示す．

t1：川端康成は「雪国」などの作品でノーベル文学賞を受賞した。
t2：川端康成は「雪国」の著者である。

図 1 テキスト対の例

```
<dataset>
  <pair id="1" label="Y">
    <t1>アテネの市域の中心にアクロポリスの丘、 北東部にリュカベットス山がそびえ、パルテノン神殿，聖イヨルイヨス礼拝堂などがある。</t1>
    <t2>パルテノン神殿の建つ丘は、アクロポリスと呼ばれている。</t2>
  </pair>
  <pair id="2" label="N">
    <t1>パルテノン神殿は、ドーリア式神殿の最高傑作と言える作品である。</t1>
    <t2>パルテノン神殿は、ヘレニズム文化の影響下で建設された。</t2>
  </pair>
  (以下省略)
</dataset>
```

図 2 BC サブタスクの学習用データ

3. 関連研究

本節では，2011 年度に NTCIR で行われた RITE タスクの研究を中心に関連研究をまとめる．3.1 項で 2011 年度の RITE において最高精度であった Pham ら[5]の手法を述べ，3.2 項で Pham らに次ぐ精度であった Akiba ら[6]の手法について述べる．

3.1. Pham らの研究

Pham らが構築した含意関係システムは以下のステップをたどって特徴量の抽出を行う．

1. 学習用データ，評価用データそれぞれについて形態素解析，品詞タグ付け，係り受け解析を行う．
2. 学習用データ，評価用データそれぞれから文章類似度などの特徴量を抽出する．

また，提供されるデータセットの文章対を Google Translate API[13]を利用して全て英語に翻訳し，同じステップをたどって特徴量を抽出した．

元の学習用データから抽出した特徴量と，英語に翻訳した後に抽出した特徴量を併せて SVM で学習を行い，判定機を構築した．評価用データに対しても同様に特徴量を抽出した後，構築した判定機による含意関係の判定を行った．

特徴量として用いたのは以下の値である．

- レーベンシュタイン距離
- Jaro-Winkler 距離
- 単語重複度
- 感情極性の比較

感情極性の比較とは，まず CaboCha を用いて文章を木構造にし，木構造のルートノードに対し単語の極性を付与する．そして文章対の文章それぞれについて極性を付与した後，極性の比較を行うことである．単語の極性は，単語感情極性対応表[12]を用いた．

最終的な判定精度として，2011 年度に行われた RITE タスク参加チームの中で最高精度の 58.0%を達成した．

3.2. Akiba らの研究

Akiba らも Pham らと同様に，学習用データセットとテスト用データセットに対してそれぞれ係り受け解析などの前処理を行い，特徴量を抽出した．また彼らは文章中に存在する述語項構造を解析し，解析結果を特徴量として用いた．述語項構造とは，述語と項との関係を表すものであり，語形成における意味的な変化は，述語項構造の変化として捉えることができる．Akiba らは，*t1* の述語項構造と *t2* の述語項構造を比較して，*t1* と *t2* に対して表 1 で示す 5 種類の関係があるか否かを特徴量とした．精度としては Pham らに次ぐ 54.8%であった．

表 1　*t1* と *t2* に対する 5 種類の関係

関係名	*t1*，*t2* の関係
Forward	*t1* が *t2* を含意し，*t2* が *t1* を含意しない
Reverse	*t2* が *t1* を含意し，*t2* が *t1* を含意しない
Bi-directional	*t1* が *t2* を含意し，*t2* が *t1* を含意する
Contradiction	*t1*，*t2* が矛盾する
Independence	*t1*，*t2* は無関係である

4. 提案手法

本節では提案手法について述べる．まず 4.1 項で提案手法の概要について述べる．次に提案手法の各フェーズでの処理について詳しく述べる．4.2 項で入力された文章に対する解析処理について述べる．4.3 項で解析された文章対からの特徴量抽出について述べる．4.4 項で述語項構造の比較により含意関係があると判断できる文章対の検出方法を述べる．4.5 項で特定の指標に基づいて含意関係がないと判断できる文章対の検出方法を述べる．4.6 項で SVM による分類について述べる．

4.1. 提案手法の概要

本研究の提案手法は，日本語文章の含意関係認識に対する一般的な既存手法である機械学習による分類を行う前に，複数の分類フェーズを加えたものである．提案手法の流れを図 3 に示す．

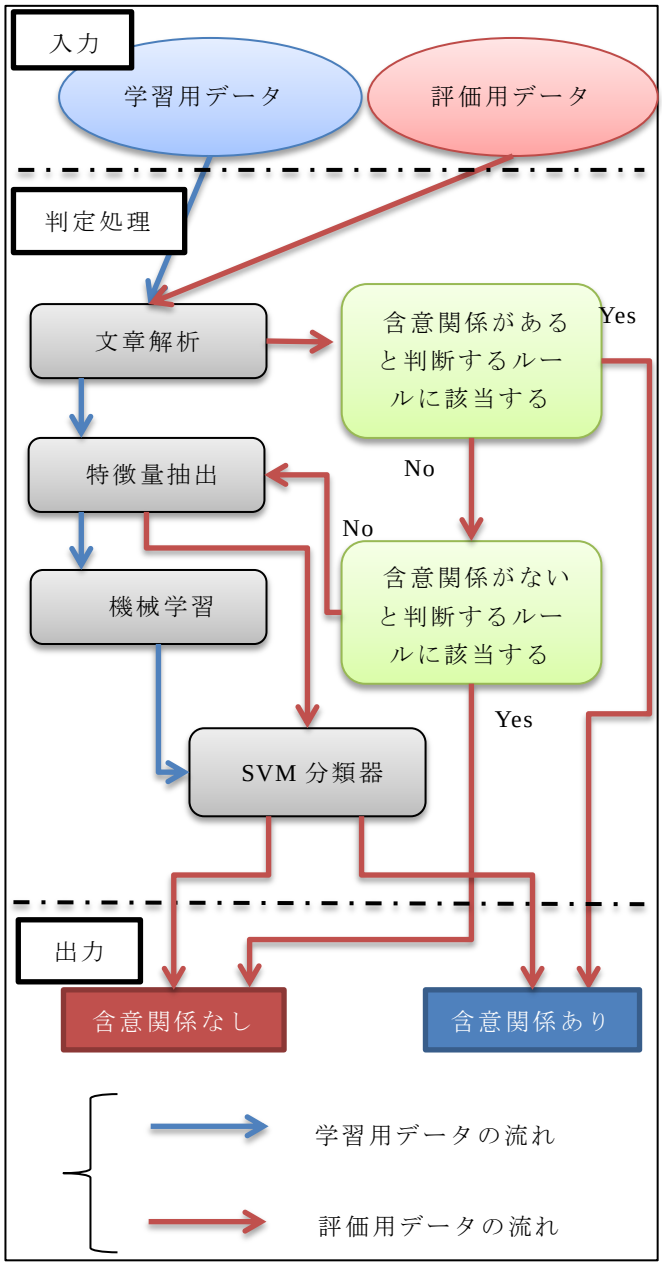

図 3 提案手法の流れ

4.2. 文章の解析処理

まず入力された文章対に対して形態素解析，品詞のタグ付け，係り受け解析を行う．また，評価データに対しては 5.2 項での述語項構造の比較を行うために述語項構造の解析を行う．提案手法で用いた文章解析ツールと解析内容を表 2 に示す．

表 2 提案手法で用いた文章解析ツールと解析内容

文章解析ツール	解析内容
MeCab[7]	形態素解析 品詞のタグ付け
CaboCha[8]	形態素解析 係り受け解析 品詞のタグ付け 固有名詞の抽出

4.3. 文章対からの特徴量抽出

解析された文章対から以下の値を特徴量として抽出し，SVM により学習する．

- 形態素の重複率

MeCab によって *t1*，*t2* を形態素に分割する．*t2* に含まれる形態素が *t1* に多く含まれている場合，含意関係がある可能性が高いと判断する．最小で 0，最大で 1 の範囲に正規化するため，*t2* の形態素の数を用いて正規化する．形態素の重複率 *morpheme_overlap* は式(1)で表される．

$$morpheme\_overlap = \frac{t1,\ t2\ で重複する形態素の数}{t2\ の形態素の数} \quad (1)$$

- 名詞の重複率

MeCab によって品詞のタグ付けがされるため，特定の品詞の抽出を行うことができる．*t2* に含まれる名詞が多く *t1* に含まれている場合，含意関係がある可能性が高いと判断する．名詞の重複率 *noun_overlap* は式(2)で表される．

$$noun\_overlap = \frac{t1,\ t2\ で重複する名詞の数}{t2\ の名詞の数} \quad (2)$$

- 文字単位の 4-gram の重複率

文字単位の n-gram の重複率のみを特徴量として精度を計測した結果，n = 4 のときに最も精度が高くなったため，文字単位の 4-gram を特徴量とした．文字単位の 4-gram の重複率 *4gram_overlap* は式(3)で表される．

$$4gram\_overlap = \frac{t1,\ t2\ で重複する\ 4-gramの数}{t2\ の\ 4-gramの数} \quad (3)$$

- 最長共通文字列の割合

t1 と *t2* どちらにも含まれる文字列の中で，最も長い文字列の長さを *lcs* とする．値を 0 から 1 までの範囲に正規化するため，*t2* の文字数を用いて正規化する．最長共通文字列の割合 *lcs_overlap* は式(4)で表される．

$$lcs\_overlap = \frac{lcs}{t2\ の文字数} \quad (4)$$

- Jaro 距離

Jaro 距離とは文章間の類似度を数値化したものである．Jaro 距離 *jaro_distance* の定義は式(5)で与えられる．

$$jaro\_distance = \begin{cases} 0\ [m = 0] \\ \frac{1}{3}\left(\frac{m}{len(t1)} + \frac{m}{len(t2)} + \frac{(m-t)}{m}\right)\ [m \neq 0] \end{cases} \quad (5)$$

式(5)において *m* とは *t2* に含まれる文字のうち出てくる順序を維持したまま *t1* にも含まれる文字の数で，*t* は *t1* にも *t2* にも含まれるが，出てくる順序が異なる文字の数である．また，*len*()は文章の文字の数を取得する関数であり，*len*(*t1*)は *t1* の文字の数，*len*(*t2*)は *t2* の文字の数である．

4.4. 含意関係があると判断できる一部の文章対検出

本項では，述語項構造の比較に基づいて含意関係があると判断できる文章対の検出方法を述べる．評価データがシステムに入力されて 4.2 項で述べた文章の解析処理がされた後，*t1*，*t2* の述語項構造を比較する．

述語項構造とは，述語と名詞の意味的関係を表すための述語と関連する項との関係を同定した構造である．日本語では，述語を中心として，述語が係り受け先となっている文節の格助詞と名詞を合わせて項とすることが多い．このときの格助詞を「格」，名詞を「名詞句」と記述する．述語項構造の例を図 4 に示す．

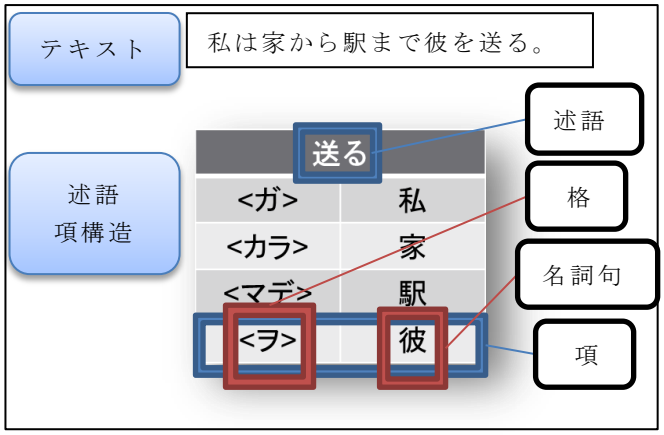

図 4 述語項構造の例

提案手法における述語項構造の作成手順を説明する．まずテキストに対して CaboCha を用いて係り受け解析を行う．次に，述語を含む文節を係り受け先とする文節を項とする．そして，項の文節を係り受け先とする文節がある場合は，項の文節と項を係り先とする

文節を結合して 1 つの項とする．図 5 に述語項構造の作成手順を示す．図 5 を用いて述語項構造の作成手順を説明する．まずテキストの係り受け関係を解析する．次に，「買う」という述語を含む文節を係り受け先とする文節「私が」，「本を」，「本屋で」をそれぞれ 1 つの項とする．さらに，項の文節「本屋で」を係り受け先とする「A 市の」という文節があるため，その 2 つの文節を結合して，「A 市の本屋で」を 1 つの項とする．

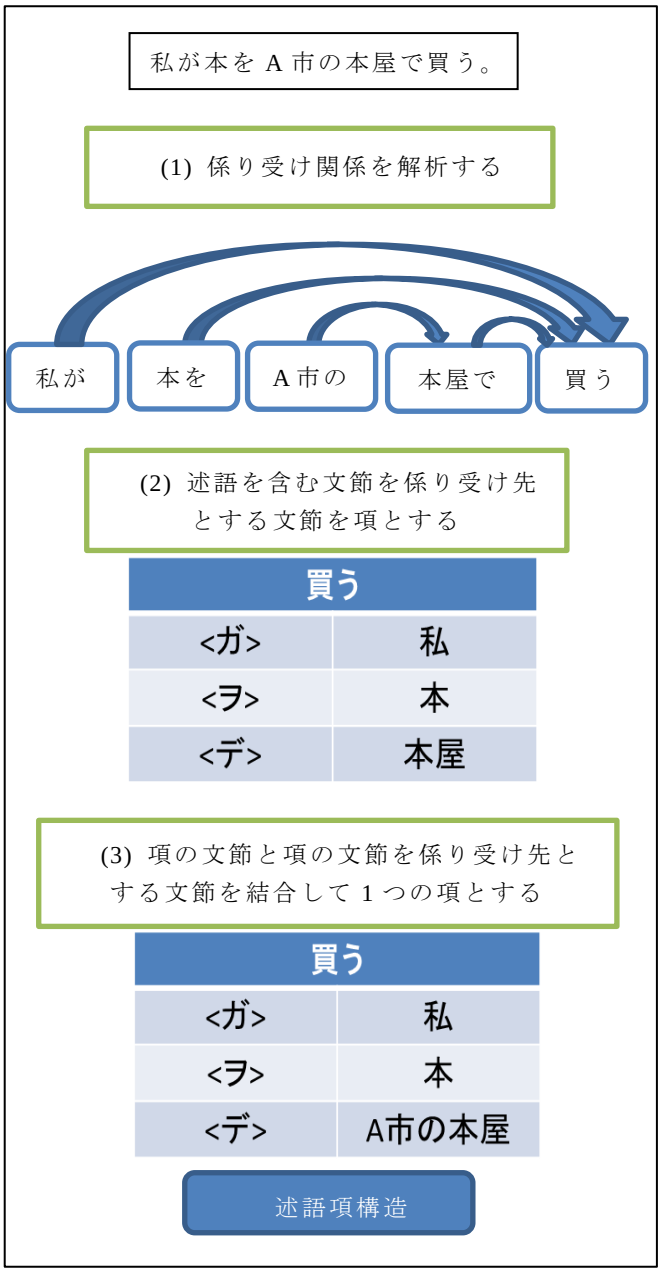

図 5 述語項構造の作成手順

t1 の述語項構造と *t2* の述語項構造を作成した後に，2 つの述語項構造の比較を行う．*t2* の述語項構造が *t1* の述語項構造に含まれているとき，*t2* は *t1* に含意されていると判定し，4.5 項，4.6 項の処理は行わない．*t2* の述語項構造が *t1* の述語項構造に含まれる条件は，t2 の全ての述語が *t1* に存在し，*t2* の述語に対する全ての項が *t1* に存在することである．

4.5. 含意関係がないと判断できる一部の文章対検出

本項では，特定の指標に基づいて含意関係がないと判断できる文章対の検出方法を述べる．人間が文章を見て含意関係がないと判断する基準の 1 つに，「*t2* にある情報が *t1* に存在しない」という条件がある．*t1* で示されていない情報が *t2* で示されていると矛盾が生じるため，*t2* は *t1* に含意されないと考えられる．しかし，一般的な表現を「情報」として捉えると，単語の言い換え表現を網羅的に取得して判断する必要があるので実現は困難である．そこで，言い換え表現がない表現，または言い換え表現がほとんど存在しない表現を用いて，含意関係がない文章対の検出を行う．

機能表現の有無による分類手法

松吉ら[9][10]は，日本語の文を構成する要素には，主に内容的な意味を表す要素（内容語）以外に，助詞や助動詞といった，主に文の構成を表す要素があるとした．松吉らは後者を総称して「機能表現」と定義した．機能表現は，助詞，助動詞，接続詞，形式名詞といった品詞の語単体からなる「機能語」と，複数の語から構成され，かつ全体として機能語のように働く表現である「複合辞」からなる．

日本語や英語などの自然言語には，多くの語に言い換えが存在する．例えば，「に関して」という表記と「に関する」という 2 つの表記を同一表現と見なすかどうかは，区分の観点により異なる．意味的機能では対応するが，文法的機能は異なるためである．松吉らは見出しを階層化することによって，多種多様な言い換えに対応した．つまり，ある階層レベルでは「に関して」と「に関する」は同一とみなすが，より下位のレベルでは別のものと見なすことにしたのである．表 3 に区分を載せる．

表 3 機能的表現の階層的見出し([9]より引用)

レベル	区分の観点	数
L1	最上位区分	341
L2	意味	435
L3	文法的機能	555
L4	機能語の交替	774
L5	音韻的変化	1187
L6	とりたて詞の挿入	1810
L7	活用	9722
L8	「です/ます」の有無	9722
L9	表記の異なり	16801

表 3 中の数とは，それぞれの区分においての機能表

現の数である．提案手法では，表 3 の中で最上位区分として「限定」として定義されている機能表現，及びそれらの変化系を用いて比較を行う．

文章に限定表現が存在する場合，その文章中にある量や範囲などを定めて，それを超えるものを認めないことを示す．*t1* と *t2* に含意関係があるとき，*t2* に限定表現が存在するならば，*t1* にも量や範囲を狭める表現が必要になる．そのため，*t2* にのみ限定表現が存在する場合，*t1* と *t2* の間に含意関係はないと判断できる．図 6 に限定表現が *t2* にのみ存在する文章の例を示す．図 6 中の下線部が限定表現を表している．

t1：女性がかかるがんの中でも 1 位を占めているのは乳癌である。

t2：乳癌は女性しかかからない。

図 6　*t2* にある限定表現が *t1* に存在しない例

図 6 の赤文字の部分が限定表現である．*t1* と *t2* に含意関係がないことがわかる．

提案手法では，松吉らが提供する日本語機能表現辞書であるつつじ[9]から，限定表現のみを対象として取り出した．

年代表現の有無による分類手法

t2 で「2005 年」や「江戸時代」といった年代表現が記述されているが *t1* には年代表現が記述されていない場合は，*t1* から *t2* の年代表現を導くことができないため含意関係がないとみなす．*t2* に年代表現があり *t1* に年代表現がない例を図 7 に示す．図 7 中の下線部が年代表現である．

t1：アンコールワットは、気軽に訪れることのできるカンボジア最大の観光地になっている。

t2：アンコールワットは１２世紀に建設された。

図 7 *t2* に年代表現があり *t1* に年代表現がない例

固有名詞の有無による分類手法

固有名詞とは人名や地名等，それ以外には存在しない特定の対象を表す名詞である．固有名詞は言い換えの表現がほとんど存在しないために，*t2* にある固有名詞が *t1* に存在しない場合には含意関係がないとみなす．*t2* にある固有名詞が *t1* に存在しない例を図 8 に示す．図 8 では *t2* の下線部「徳川家康」という固有名詞が *t1* に存在しない．

t1：清浄院は加藤清正の継室である。

t2：清浄院は，徳川家康の養女である。

図 8　*t2* にある固有名詞が *t1* に存在しない例

4.6. SVM による分類

4.4 項で述語項構造に含意関係があると判断されず，4.5 項で含意関係がない文章対の検出ルールに該当しなかった文章対は SVM 分類器によって含意関係を判断する．4.3 項で述べた特徴量を使用し，学習には Chang らが開発した LIBSVM[14]を使用する．SVM のカーネルには線形カーネルを用いた．

5. 評価実験

本節では提案手法による含意関係認識の評価実験とその結果について述べる．

5.1. 使用データ

2011 年度に行われた RITE を RITE1，2012 年度に行われた RITE を RITE2 と記述する．実験には RITE1 と RITE2 の BC サブタスクのデータのうち，含意関係が示されているデータを用いた．*t1*，*t2* からなるテキスト対を 1 問と表すと，RITE1 の学習用データは 500 問，RITE1 の評価用データは 500 問，RITE2 の学習用データは 611 問であり，計 1,611 問のデータを使用した．このうち含意関係があるテキスト対が 740 問，含意関係がないテキスト対が 871 問であった．5.4 項の Pham らの手法との比較実験においては，RITE1 のデータを用いて，学習用に 500 問，評価用に 500 問を使用した．

5.2. 一部の文章対について含意関係があると判断するルールに対する実験

4.4 項で述べたように，*t2* の述語項構造が *t1* の述語項構造に含意される場合は含意関係があると判定する．述語項構造に含意関係がある問題数を検出問題数とし，全データ 1,611 問を対象としたときの検出問題数，検出問題の中で含意関係がある問題数，精度及び再現率を計測した結果を表 4 に示す．表 4 における精度とは，検出問題数に対する含意関係がある問題数の割合であり，式(6)で表される．

$$精度 = \frac{含意関係がある問題数}{検出問題数} \qquad (6)$$

表 4　述語項構造の比較による検出問題数と精度

検出問題数	含意関係がある問題数	精度(%)	再現率(%)
19	16	84.2	2.2

5.3. 一部の文章対について含意関係がないと判断するルールに対する実験

4.5 項で述べたように，特定の指標に基づいて含意関係がないと判断された文章対に対し，精度の確認実験を行う．5.2 項と同様に全データ 1,611 問を対象とした時の検出問題数，検出問題の中で含意関係がない問

題数，精度及び再現率を計測した．結果を表 5 に示す．表 5 における精度とは，検出問題数に対する含意関係がある問題数の割合であり，式(7)で表される．

$$精度 = \frac{含意関係がない問題数}{検出問題数} \quad (7)$$

表 5 ルールごとの検出問題数と精度

	検出問題数	含意関係がない問題数	精度(%)	再現率(%)
機能表現の有無	10	9	90.0	1.0
年代表現の有無	12	11	91.7	1.3
固有名詞の有無	153	135	88.2	15.5

5.4. 総合実験

提案手法は 3 つの分類フェーズを持ち，それぞれのフェーズを次のように定義する．

述語項構造の含意関係による分類

Phase1 含意関係があると判断できる一部の文章対検出

Phase2 含意関係がないと判断できる一部の文章対検出

Phase3 SVM による分類

RITE1 のデータを用いたときの提案手法による各分類フェーズで処理した問題数，正解数及び精度を表 6 に示す．また，提案手法と RITE1 の BC サブタスクで最高精度を出した Pham らの手法[5]の正解数，問題数と精度を表 7 に示す．表 7 より，提案手法では Pham らの手法より 2.8%精度が向上したことがわかる．表 6 より，Phase1，Phase2 における精度が 80%を超えている．含意関係があると判断できる一部の文章対と，含意関係がないと判断できる一部の文章対を高精度で検出したことで，全体としての含意関係認識の精度が向上したことが確認できる．

表 6 各分類フェーズの問題数，正解数，精度

		問題数	正解数	精度(%)
分類フェーズ	Phase1	6	5	83.3
	Phase2	28	23	82.1
	Phase3	466	276	59.2

表 7 提案手法と Pham らの手法での問題数，正解数，精度

	問題数	正解数	精度(%)
提案手法	500	304	60.8
Pham らの手法	500	290	58.0

6. まとめ

本稿では，自然言語の意味解析における研究分野の一つである含意関係認識の高精度な分類手法を提案した．自然言語の意味解析における問題として，プログラミング言語などの人工言語に比べると，文法が無秩序であり，言い換えや比喩が数多く存在するため，言語の意味を解析する精度が良くないことがあった．提案手法では，ある指標によって含意関係認識が容易な文章対が入力として与えられた場合には，その指標に基づいて判定を行い，含意関係認識が難しいと考えられる文章対に対しては機械学習によって構築した判定機を用いて判定することによって，全体の精度向上を果たした．機械学習以外の分類方法として，述語項構造の比較による分類，文章中の機能表現を考慮した分類，含意関係がない文章対の検出ルールによる分類の 3 つを加えた．実験の結果，2011 年度 RITE の BC サブタスクで最高精度であった Pham らの精度 58.0%を超える 60.8%の精度を出すことができた．

提案手法では，機械学習を用いた分類を行う前に，含意関係があると判断できる一部の文章対検出と，含意関係がないと判断できる一部の文章対検出を 80%以上と高い精度で行うことにより，全体の精度を向上させることができた．しかし，使用した全データにおいて，含意関係があると判断できる文章対検出の再現率は 2.2%，含意関係がないと判断できる文章対検出の再現率は 17.8%であり，再現率が低い．今後の課題として，これらの再現率の向上があり，含意関係がある文章対検出ルールと含意関係がない文章対検出ルールの増加を検討している．

謝辞

本研究で用いた文章データは，NTCIR の提供によるものである．ここに感謝致します．

参 考 文 献

[1] PASCAL 2 , “http://www.pascal-network.org/” , (2013/01/07 アクセス)

[2] Text Analysis Conference, “http://www.nist.gov/tac/” , (2013/01/07 アクセス)

[3] NTCIR, “http://research.nii.ac.jp/ntcir/index-ja.html”, (2013/01/07 アクセス)

[4] RITE-2, “http://www.cl.ecei.tohoku.ac.jp/rite2/doku.php?id=wiki:%E3%83%A1%E3%82%A4%E3%83%B3%E3%83%9A%E3%83%BC%E3%82%B8” , (2013/01/07 アクセス)

[5] Quang Nhat Minh Pham, Le minh Nguyen, Akira Shimazu, “A Machine Learning based Textual Entailment Recognition System of JAIST Team for NTCIR9 RITE”, Proceedings of NTCIR-9 Workshop Meeting, December 6-9, 2011, Tokyo, Japan, 2011.

[6] Yasuhiro Akiba, Hirotoshi Taira, Sanae Fujita, Kaname Kasaharam, and Masaaki Nagata, “NTTCS

Textuak Entailment Recognition System for NTCIR-9 RITE", In Proc. of NTCIR-9 Workshop Meeting, pp.330-334, 2011.

[7] MeCab: Yet Another Japanese Dependency Structure Analyzer, "https://code.google.com/p/mecab/" , (2013/01/07 アクセス)

[8] CaboCha/南瓜: Yet Another Japanese Dependency Structure, "http://code.google.com/p/cabocha/" , (2012/12/28 アクセス)

[9] つつじ:日本語機能表現辞書, "http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji/", (2013/01/07 アクセス)

[10] 松吉俊, 佐藤理史, 宇津呂武, "日本語機能表現辞書の編纂", 自然言語処理, Vol.14, No.5, pp123-146, 2007.

[11] 高村大也, 乾孝司, 奥村学, "スピンモデルによる単語の感情極性抽出", 情報処理学会論文誌ジャーナル, Vol.47 No.02 pp. 627--637, 2006.

[12] 単語感情極性対応表, " http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html " , (2013/01/07 アクセス)

[13] Google Translate API, "https://developers.google.com/translate/?hl=ja" , (2013/01/07 アクセス)

[14] LIBSVM -- A Library for Support Vector Machines, "http://www.csie.ntu.edu.tw/~cjlin/libsvm/" , (2013/01/07 アクセス)