

## 日本語テキスト含意認識における意味構造の利用

砂川 健一<sup>†</sup> 石塚 満<sup>‡</sup>東京大学工学部電子情報工学科<sup>†</sup>東京大学大学院情報理工学系研究科<sup>‡</sup>

## 1. はじめに

テキスト含意認識 (Recognizing Textual Entailment, RTE) とは自然言語処理のタスクの一つである。これは二つの文、T (Text) と H (Hypothesis) が与えられた時に、T が H を含意するか、つまり文 T から文 H が推論可能かどうかを判定する。このように言語の意味を推論する技術が確立されればよりインテリジェントなテキスト処理が可能となるため、多くの分野での応用が期待される基礎タスクとして位置づけられている。

本研究では単語の意味的類似度および構文解析結果から意味構造を利用した含意判定の特徴量を計算する。それに際し、一般の自然言語処理タスクにおいてもそうであるが、日本語でのテキスト含意認識は英語と比べて、分かち書きがされていない、省略が多い、格構造がはっきりしにくい等の異なる点が数多く存在する。本研究では英語には存在しない「助詞」を利用して含意判定の特徴量を算出する手法を用い、判定精度を高めることを目指す。

## 2. 本システムで利用する言語情報

## 2.1 概念辞書に基づく単語間意味的類似度

概念辞書では単語が synset と呼ばれる同義語のグループに分類され、他の同義語のグループとの関係が記述されている。この情報に基づいて単語間意味的類似度を定量的に算出する手法が多数提案されており、本研究では概念辞書である日本語 WordNet を用いて次の3種類を使用した：

Semantic Structure for Japanese Textual Entailment Recognition

<sup>†</sup> K. Sunagawa, Univ. of Tokyo

<sup>‡</sup> M. Ishizuka, Univ. of Tokyo

1: Path Similarity

2: Lch (Leacock-Chodorow) Similarity

3: Wup (Wu-Palmer) Similarity

但し、本研究では上の意味的類似度をより含意認識に適したものにした「含意関係度」を特徴量及び特徴量の計算に使用している [yokote 11]。

## 2.2 形態素解析器、構文解析器による文節係り受け情報

構文解析はテキスト含意認識において、意味構造を明示するという点で有効である。本研究では形態素解析器に JUMAN、構文解析器に KNP を用いたデータを使用した。

## 3. 意味構造を加味した特徴量の生成

## 3.1 単語の重要度による単語関係間意味的類似度

以下の例文を考える。

T: 彼は銭湯に行き私はスーパーに行った。

H: 私は銭湯に行き彼はスーパーに行った。

この例では、形態素のオーバーラップでみると完全に一致し、係り受け関係でも一致するが、T は H を含意しない。この場合、H では「私」と「行く」の関係よりも「私」と「銭湯」の関係が含意判定に影響するであろうと考えるのが普通である。このような場合の単語関係間の意味的類似度の抽出に対応するため、「行く」のような重要度が低い単語の影響を小さくし、「行く」にかかっている語同士の関係の影響を大きくするように特徴量を設定した。なお、重要度の指標として、文書全体の各単語の出現頻度 (TF) の逆数を取った。

### 3.2 助詞（一部助動詞）による単語関係間類似度、格関係類似度

TとHの格関係（主格・目的格など）は含意判定に大きく影響を与える要素だが、ただ主格を与える助詞をとっても、「が」と「は」とでは係り受け解析で切り捨てられてしまうかもしれない意味構造に対する役割が異なる。本研究では主要な助詞について単語関係間類似度、各関係類似度に対する貢献度を算出し、含意判定の特徴量に組み込んだ。

#### 謝辞

本研究において多くのアドバイスを頂いた石川開様、土田正明様をはじめNEC 情報・メディアプロセッシング研究所の皆様に深く感謝致します。

#### 参考文献

[yookote 11]K. Yokote, S. Tanaka, D. Bollegala and M. Ishizuka:Useful Conceptual Semantic Information for Textual Entailment Recognition