

令和 2 年 6 月 5 日現在

機関番号：13901

研究種目：若手研究(B)

研究期間：2016～2019

課題番号：16K16110

研究課題名（和文）大規模テキストからの語横断的フレーム知識の獲得と意味解析への応用

研究課題名（英文）Acquisition of cross-word frame knowledge from large texts and its application to semantic analysis

研究代表者

笹野 遼平（SASANO, RYOHEI）

名古屋大学・情報学研究科・准教授

研究者番号：70603918

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究課題では、自動獲得されたフレーム知識である京都大学格フレームを、人手で整備されたフレーム知識であるFrameNetに対応付けることにより、語横断的フレーム知識を構築し、さらに対応付け知識に基づく日本語意味役割付与システムを構築した。また、異なる語彙、異なる言語間のフレームの対応付けの高精度化に向けて、特定の意味クラスに属する単語ベクトルが、意味ベクトル空間においてどのように分布しているかの分析を行った。

研究成果の学術的意義や社会的意義

本研究では、自動獲得されたフレーム知識を、人手で整備されたフレーム知識に対応付けることで、語横断的な、言語横断的なフレーム知識を構築できることを示した。さらに、構築した知識を用いることで、英語を対象に構築された意味役割タグ付きコーパスから、日本語意味役割付与システムを構築できることを示した。また、特定の意味クラスに属する単語ベクトルが意味ベクトル空間においてどのように分布しているかに関する分析は、単語ベクトルを用いたシステムの構築において有用な知見となると期待できる。

研究成果の概要（英文）：In this research project, we constructed cross-word frame knowledge by linking the Kyoto University case frames, which is automatically acquired frame knowledge, to the manually-crafted frame knowledge FrameNet and developed a Japanese semantic role labeling system based on the constructed frame knowledge. In addition, we investigated the distribution of words that belong to a certain word class in a pre-trained word vector space to improve the accuracy of frame mapping between different lexicons and different languages.

研究分野：自然言語処理

キーワード：格フレーム 意味役割付与 単語ベクトル

1. 研究開始当初の背景

自然言語テキストの意味を計算機で扱うための基盤技術の 1 つである項構造解析において、人手で整備された意味役割付きフレームと、大規模コーパスから自動獲得したフレーム知識の 2 種類のフレーム知識が利用されてきた。しかし、人手で整備されたフレームは、人間の直感的理解に近い文の意味解析が可能であるという特長があるものの、大規模化のコストが非常に大きいという問題があり、一方、自動獲得したフレーム知識は、カバレッジは大きいものの、粒度が人間の直感的理解と一致しているとは限らず、また、語ごとに個別に構築されており、類似したフレームの関係を捉えられないという問題があった。

2. 研究の目的

自動構築されたフレーム知識と人手で整備された意味役割付きフレーム知識を、単語意味ベクトルを用いて関係付けることで、カバレッジが大きく、かつ、意味役割情報を含む大規模なフレーム知識を構築することを目的とする。特に、近年急速に発展している言語横断的な単語意味ベクトル獲得技術を活用し、異なる言語間のフレーム知識を結びつけることで言語横断的なフレーム知識の構築を行い、構築した知識に基づく意味役割システムを実現する。また、フレームの対応付け精度の向上のため、特定の意味クラスに属する単語ベクトルが、意味ベクトル空間においてどのように分布しているか明らかにする。

3. 研究の方法

約 70 億文の Web テキストから自動構築された日本語格フレームである京都大学格フレームを、大規模コーパスから事前に学習した単語意味ベクトルを利用し、英語を対象に人手で整備されたフレーム知識である FrameNet、さらにその日本語版である日本語 FrameNet にそれぞれ対応付けることにより、語横断的なフレーム知識の構築を行った。さらに、構築した知識に基づく日本語を対象とした意味役割付与システムを実現した。

また、特定の意味クラスに属する単語ベクトルが、意味ベクトル空間においてどのように分布しているか明らかにするため、ある動詞の直接目的語となる単語集合、および、シソーラスで共通の語を上位語として持つ単語集合を対象に、それらの単語の意味ベクトルの分布のモデル化を行った。

4. 研究成果

(1) 自動構築した日本語格フレームと英語 FrameNet の対応付け

日本語を対象に自動獲得されたフレーム知識である京都大学格フレームを、英語を対象に人手で整備されたフレーム知識である FrameNet に対応付けることにより、語横断的なフレーム知識を構築した。さらに、対応付け知識に基づく日本語意味役割付与システムの構築に取り組んだ。

フレームの対応付けに関しては、まず対訳辞書に基づく教師あり手法により日本語と英語間の言語横断単語意味ベクトルを構築し、続いて構築したベクトルを用い、京大格フレームに含まれる用例情報と、

FrameNet に含まれる注釈付き文から得られる用例情報の類似度を算出することで、京大格フレームに含まれるフレームを FrameNet のフレームに対応付けた。

意味役割付与システムの構築に関しては、与えられた日本語テキストを、京大格フレームに基づく述語項構造解析器である KNP を用いて解析することで、入力されたテキスト中の要素を京大格フレームの格に対応付け、その対応付け結果を FrameNet の意味役割に変換することにより、日本語テキストを対象とした意味役割付与システムを構築した(図 1)。評価は日本語版 FrameNet に含まれる注釈文を用いて行い、日本語の教師データを使わない設定において、39.2%の精度でフレームを同定、28.8%の精度でフレームと意味役割のペアを正しく推定できることを示した。

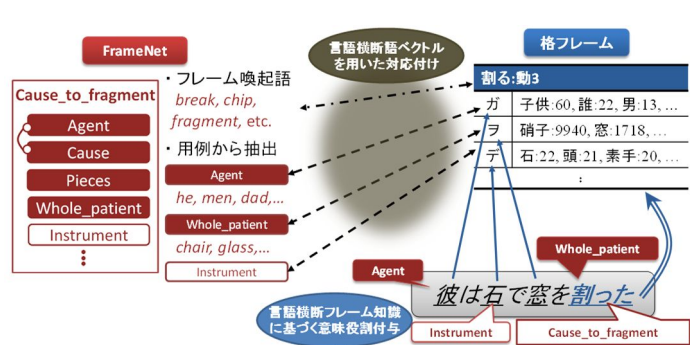


図 1 対応付け情報を用いた日本語意味役割付与

## (2) 自動構築した日本語格フレームと日本語 FrameNet の対応付け

日本語を対象に自動獲得されたフレーム知識である京都大学格フレームを、日本語を対象に人手で整備されたフレーム知識である日本語 FrameNet に対応付けることにより、語横断的フレーム知識を構築した。具体的には、動詞ごとに、その動詞の各格フレームを、その述語を喚起語として含む日本語 FrameNet フレームの 1 つに対応づけた。対応付けは格フレームの格に記述されている用例と、日本語 FrameNet の用例の類似度を用いて行った。

対応先の候補となる JFN フレームが 2 つのみである 19 動詞 119 フレームを対象とした実験の結果、約 79%の精度で正しいフレームに対応づけることができることを示した。また、与えられた述語を含む用例を重視する場合とそうでない場合を比較した結果、与えられた述語を含む用例を重視することで精度が向上することを確認した。

## (3) 単語ベクトル空間における単語クラス分布の分析

フレームの対応付け精度の向上のため、事前学習された単語ベクトル空間において、特定の単語クラスに属する単語ベクトルの分布の分析を行った。具体的には、分布についていくつかの仮定をおき、各仮定に基づいた分布をモデル化(図 2)を行い、クラス境界検出および典型性評価の 2 タスクにおける各モデルの精度を比較することにより分布の分析を行った。

選択的選好と WordNet のデータセットを用いた実験の結果、いずれのタスクにおいても、与えられた単語ベクトル集合(正例)の Centroid からの距離のみを考慮したモデルでは十分な精度が得られず、高い精度の実現のためには、対象の単語クラスに属さない単語の意味ベクトル(負例)の分布も考慮することが必要であることを明らかにした。また、クラス境界検出においては識別学習に基づくモデルが優れているのに対し、典型性評価においてはオフセットに基づくモデルが優れていることを示した。

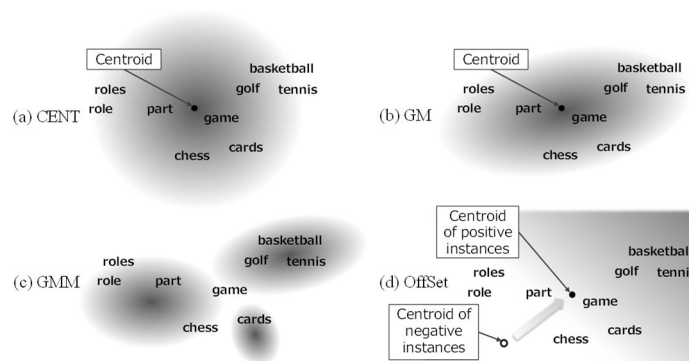


図 2 単語クラスのモデル化の例

## 5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 山田康輔, 笹野遼平, 武田浩一	4. 巻 35
2. 論文標題 「いいね」「シェア」をした投稿のテキスト情報を利用したSNSユーザの性格推定	5. 発行年 2020年
3. 雑誌名 人工知能学会論文誌	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 笹野 遼平、奥村 学	4. 巻 24
2. 論文標題 大規模コーパスに基づく日本語二重目的語構文の基本語順の分析	5. 発行年 2017年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 687 ~ 703
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.24.687	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計6件（うち招待講演 0件/うち国際学会 3件）

1. 発表者名 Ryohei Sasano and Anna Korhonen
2. 発表標題 Investigating Word-Class Distributions in Word Vector Spaces
3. 学会等名 58th Annual Meeting of the Association for Computational Linguistics (国際学会)
4. 発表年 2020年

1. 発表者名 大柳慶悟, 笹野遼平, 武田浩一
2. 発表標題 自動構築された格フレームと日本語FrameNetの対応付け
3. 学会等名 言語処理学会 第26回年次大会
4. 発表年 2020年

1. 発表者名 Kosuke Yamada, Ryohei Sasano, Koichi Takeda
2. 発表標題 Incorporating Textual Information on User Behavior for Personality Prediction
3. 学会等名 ACL 2019 Student Research Workshop (国際学会)
4. 発表年 2019年

1. 発表者名 笹野遼平
2. 発表標題 自動構築した日本語格フレームと英語FrameNetの対応付け
3. 学会等名 言語処理学会 第25回年次大会
4. 発表年 2019年

1. 発表者名 高橋良輔, 笹野遼平, 武田浩一
2. 発表標題 格要素の典型性と連想性に基づいた換喩判定
3. 学会等名 言語処理学会 第25回年次大会
4. 発表年 2019年

1. 発表者名 Ryohei Sasano and Manabu Okumura
2. 発表標題 A Corpus-Based Analysis of Canonical Word Order of Japanese Double Object Constructions
3. 学会等名 54th Annual Meeting of the Association for Computational Linguistics (国際学会)
4. 発表年 2016年

〔図書〕 計1件

1. 著者名 笹野 遼平、飯田 龍	4. 発行年 2017年
2. 出版社 コロナ社	5. 総ページ数 196
3. 書名 文脈解析 -述語項構造・照応・談話構造の解析-	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----