

令和 3 年 6 月 15 日現在

機関番号：32689

研究種目：基盤研究(B) (一般)

研究期間：2018～2020

課題番号：18H03286

研究課題名(和文) 言内・言外の意味を統合した知識フレームの獲得と言語理解への応用

研究課題名(英文) Acquisition of Knowledge Frame with Denotational and Connotational Meanings and its Application to Text Understanding

研究代表者

河原 大輔 (Kawahara, Daisuke)

早稲田大学・理工学術院・教授

研究者番号：10450694

交付決定額(研究期間全体)：(直接経費) 13,300,000円

研究成果の概要(和文)：計算機による自然言語理解の実現に向けて、述語項構造を基本とする格フレームや「イベント」に対して、言内の意味(denotation)と言外の意味(connotation)に関する知識を獲得した。denotationに関する知識は、深層学習技術を利用して格フレームと意味フレーム(FrameNet)の対応付け、および、意味フレームの自動推定を行った。connotationに関する知識は、イベントに対する感情知識を漸進的に獲得した。また、獲得した知識を深層学習モデルにおいて利用する手法についても考案した。

研究成果の学術的意義や社会的意義

従来の自動構築された格フレームは言内・言外ともに意味知識をもっておらず、統語的な言い換えは認識できても、意味的な同義関係は認識できなかった。本研究の成果はこれらの知識を獲得したもので、計算機による自然言語の意味理解に一步踏み出すことができたと考える。今後の展開として、本研究の成果を利用することによって情報アクセス技術を高度化し、情報の取得や利活用といった人間の基本的な知的活動を強力に支援することができるようになることを期待できる。

研究成果の概要(英文)：We have acquired knowledge about denotation and connotation for case frames and "events" based on predicate-argument structures to achieve natural language understanding. For denotation knowledge, we mapped case frames to semantic frames of FrameNet and induced semantic frames using deep learning techniques. For connotation knowledge, we gradually acquired emotion knowledge for events. We also devised a method to use knowledge in deep learning models.

研究分野：自然言語処理

キーワード：言語理解 知識フレーム 言内の意味 言外の意味 述語項構造

### 1. 研究開始当初の背景

これまで、計算機による自然言語理解を目標として、述語項構造解析のような基盤解析技術が活発に研究されてきた。述語項構造解析は、テキスト中における「誰がどこでいつ何をした」のような 5W1H を同定する技術である。この解析は、述語項構造ごとに集約・整理した知識源である格フレームに基づいて行われる。しかし、これまでの格フレームに基づく述語項構造解析は格の正規化などの統語的解釈に終始しており、述語や項の異なる 2 つの表現が同じ意味をもつかどうかを判定することはできていない。この問題を解決するには、格フレームに対して、言内つまり直示的な意味(denotation)、および、言外の意味(connotation)に関する知識を与える必要がある。

### 2. 研究の目的

計算機による自然言語理解を実現するためには、統語的な解釈を越えて、意味の領域に踏み込む必要がある。文・文章の意味の根幹は述語にあり、これを細かい粒度でとらえた知識を計算機に与える必要がある。これらの知識として、本研究では denotation に関する知識、および、connotation に関する知識の 2 つを対象とし、これらを自動獲得することを目的とする。本研究の成果によって、自然言語理解に端緒が開け、情報アクセス技術の飛躍的な高度化がもたらされ、情報の取得や利活用といった人間の基本的な知的活動を強力に支援することができるようにすると期待できる。

### 3. 研究の方法

denotation に関する知識については、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大格フレームと、英語版 FrameNet、および、日本語版 FrameNet との対応付けに取り組み、これまで独立したフレームとして扱われていた格フレーム間の意味的關係に関する知識の獲得を目指した。また、京大格フレームは粒度が細かすぎる等の問題があることが判明したため、新たなフレーム知識として、BERT 等に代表される文脈化単語ベクトルを用いた意味フレーム構築に取り組んだ。

connotation に関する知識については、感情知識にフォーカスし、ブートストラップ的手法を用いることによって、上記 Web テキストから漸進的に感情知識を獲得した。また、connotation 知識の利用に関して、感情知識を深層学習モデルにどのようにして組み込むかについて研究した。近年の自然言語処理で一般的に用いられている seq2seq の枠組みにおいて、生成タスクに感情分類タスクを融合する手法を考案し実験を行った。

### 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

日本語と英語間の言語横断単語意味ベクトルを用い、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、京大格フレーム中の各フレームを FrameNet に対応付けた(図 1)。日本語版 FrameNet の注釈文を用いた評価実験を行い、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。

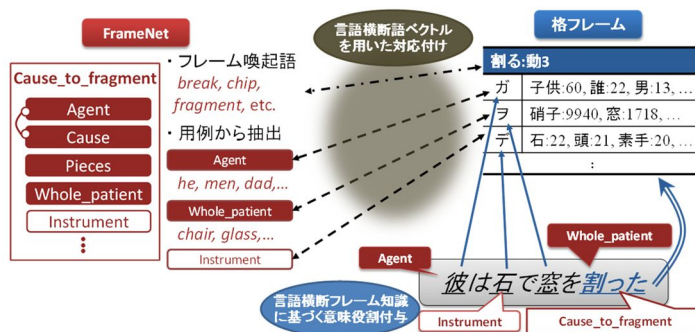


図 1 対応付け情報を用いた日本語意味役割付与

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

京大格フレームに含まれる用例情報と、日本語版 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、京大格フレーム中の各フレームを FrameNet に対応付けた。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文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文例を重視することで精度が向上することを示した。

### (3) フレーム知識の自動獲得に向けた文脈化単語埋め込みの有用性の検証

人手で整備されたフレーム知識リソースにおいて 2 種類以上のフレームを喚起する動詞に着目し、それらの動詞が喚起するフレームの違いを ELMo や BERT に代表される文脈化単語埋め込みがどのくらい捉えているかを、各用例の文脈化単語埋め込みのクラスタリング結果とそれらに付与されたフレームを比較することにより調査した。実験は FrameNet と PropBank の 2 つのフレーム知識リソースを対象に、7 種類の文脈化単語埋め込みを用いて行い、BERT や XLNet、ALBERT による文脈化単語埋め込みが、人手で付与されたフレームの違いを高い精度で捉えていることを示した。

### (4) マスクされた単語の埋め込みと 2 段階クラスタリングを用いた動詞の意味フレーム推定

テキスト中の動詞が喚起する意味フレームの推定において、マスクされた単語の埋め込みと 2 段階クラスタリングを利用する手法を提案した(図 2)。具体的には、動詞のフレーム推定に利用する埋め込みとして、従来手法で使用された推定対象動詞の文脈化単語埋め込みに加え、その動詞をマスクした場合の文脈化単語埋め込みを利用した。また、1 段階目で動詞ごとに用例のクラスタリングを行った後、2 段階目で動詞横断的なクラスタリングを行うという 2 段階クラスタリングを利用した。FrameNet を用いた評価実験の結果、従来手法より高い精度を達成すること、マスクされた単語の埋め込みおよび 2 段階クラスタリングが有効であることを示した。

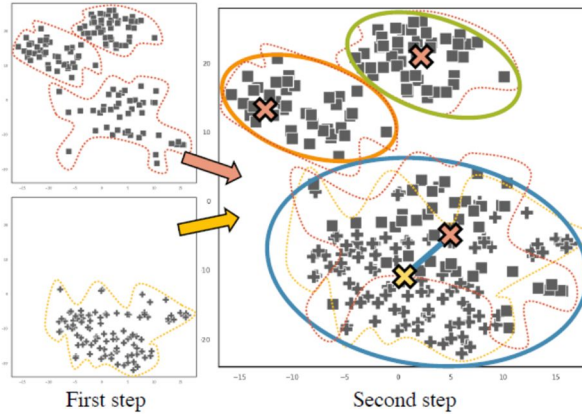


図 2 2 段階クラスタリングの流れ

### (5) イベントに対する感情知識の獲得

connotation 知識として感情知識にフォーカスした。感情を付与する単位を検討した結果、格フレームではなく、それを構成する個々の述語項構造(イベント)とすることにした。これは、格フレームには類似した項がクラスタリングによってまとまっているが、それぞれの項によってとりうる感情が異なる場合が一定数あり、無視できないためである。イベントに対する感情知識を獲得するために、ブートストラップ法によって漸進的に知識獲得を行う手法を提案した。本手法は、まず大規模日本語コーパスから述語項構造で表現されるイベントのペアを抽出する。このイベントのペアは係り受け関係にあるものとし、係り元イベントを前件イベント、係り先イベントを後件イベントと呼ぶ。次に、数十個のシード単語(形容詞)を用いて、例えば「試験に合格して、とてもうれしい」というペアから「試験に合格する」はポジティブ、「自転車でこけて、悲しい」というペアから「自転車でこける」はネガティブという感情タグ付きイベントコーパスを得る。これを用いて汎用文脈言語モデル BERT に基づくイベント感情分類器を構築し、抽出済みイベントペアの前件イベントに対して感情を推定する。この推定結果の中から分類器が自信をもって分類した感情付きイベントを訓練データに追加することができる(図 3 の High-confidence eventA-sentiment)が、これは通常のブートストラップ手法である。本研究では、イベントペアは同じ感情をもちやすいことを利用して、もう一つのブートストラップ手法を提案した。これは、感情を付与した前件イベントとペアになっている後件イベントを抽出し、閾値以上の頻度で出現する後件イベントを獲得する。これらの後件イベント集合に前件イベントに付与した感情を付与し、さらにこれらの後件イベントとペアになっている前件イベントを抽出する(図 3 の Event pair extension eventA-sentiment)。このデータを新たな訓練データとして加える。この 2 種類のブートストラップ手法を用いることによって、イベントに対する感情推定精度 88.2% を達成した。漸進的に獲得した感情付きイベント集合、またイベント感情推定モデルが成果である。

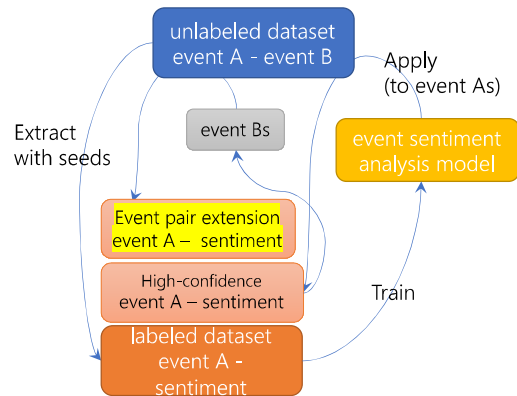


図 3 イベント感情タグ付きデータ抽出の流れ

#### (6) 感情を考慮した対話システムの開発

connotation 知識の利用に焦点を当て、感情知識を深層学習モデルにどのようにして組み込むかについて研究した。近年の自然言語処理で一般的に用いられている seq2seq の枠組みにおいて、生成タスクに感情分類タスクを融合することによって、感情知識を考慮した生成を行う手法を提案した。具体的には、seq2seq 事前学習モデル BART をベースモデルとし、感情分類タスクと対話応答生成タスクを融合したモデルを構築した。生成した応答について自動および人手評価を行ったところ、感情を考慮した生成ができており、また流暢さや発話に対する関連性も同時に向上することを確認した。本手法は、生成・分類を含め様々なタスクに応用することができ、感情知識を深層学習に組み込む手法として今後の発展が期待できる。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計13件（うち招待講演 0件 / うち国際学会 5件）

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 Qianying Liu, Wenyu Guan, Sujian Li and Daisuke Kawahara
2. 発表標題 Tree-structured Decoding for Solving Math Word Problems
3. 学会等名 In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP2019) (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 山田康輔, 笹野遼平, 武田浩一
2. 発表標題 フレーム知識の自動獲得に向けた文脈化単語埋め込みの有用性の検証
3. 学会等名 情報処理学会 第244回自然言語処理研究会
4. 発表年 2020年

1. 発表者名 高橋良輔, 笹野遼平, 武田浩一
2. 発表標題 文脈化単語埋め込みを用いた慣用句判定
3. 学会等名 情報処理学会 第245回自然言語処理研究会
4. 発表年 2020年

1. 発表者名 山田康輔, 笹野遼平, 武田浩一
2. 発表標題 マスクされた単語の埋め込みと2段階クラスタリングを用いた動詞の意味フレーム推定
3. 学会等名 言語処理学会 第27回年次大会
4. 発表年 2021年

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

1. 発表者名 Kosuke Yamada, Ryohei Sasano, Koichi Takeda
2. 発表標題 Verb Sense Clustering Using Contextualized Word Representations for Semantic Frame Induction
3. 学会等名 In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 Kosuke Yamada, Ryohei Sasano, Koichi Takeda
2. 発表標題 Semantic Frame Induction using Masked Word Embeddings and Two-Step Clustering
3. 学会等名 In Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (国際学会)
4. 発表年 2021年

1. 発表者名 井手竜也, 河原大輔
2. 発表標題 生成と分類のマルチタスク学習による感情が考慮された対話応答生成
3. 学会等名 言語処理学会 第27回年次大会
4. 発表年 2021年

1. 発表者名 榮田亮真, 河原大輔
2. 発表標題 逆翻訳とフィルタリングによる擬似対話コーパスの生成とそれを用いた対話システムの学習
3. 学会等名 言語処理学会 第27回年次大会
4. 発表年 2021年

1. 発表者名 栗原健太郎, 河原大輔
2. 発表標題 ファクトチェック支援のための含意関係認識システム
3. 学会等名 言語処理学会 第27回年次大会
4. 発表年 2021年

1. 発表者名 Tatsuya Ide, Daisuke Kawahara
2. 発表標題 Multi-Task Learning of Generation and Classification for Emotion-Aware Dialogue Response Generation
3. 学会等名 NAACL Student Research Workshop (SRW) 2021 (国際学会)
4. 発表年 2021年

〔図書〕 計0件

〔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
研究分担者	笹野 遼平 (Sasano Ryohei) (70603918)	名古屋大学・情報学研究科・准教授  (13901)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------