

特許中の画像とテキストを用いた手順オントロジーの構築

Construction of a Procedural Ontology Using Images and Texts in Patents

中央大学理工学部ビジネスデータサイエンス学科教授

難波 英嗣

2001年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了。博士（情報科学）。東京工業大学精密工学研究所助手、広島市立大学大学院情報科学研究科准教授等を経て、2019年より中央大学理工学部教授。自然言語処理、テキストマイニングの研究に従事。

✉ nanba@kc.chuo-u.ac.jp

☎ 03-3817-1883

1 はじめに

ある特定の目的を達成するための一連の手続きを記したものを手順テキストと呼ぶとき、類似の手順テキストの集合から抽出された典型的な手順が手順オントロジーである。本研究では、この手順オントロジーを構築する。

一般にオントロジーの人手による構築は非常にコストがかかる。このため、自然言語処理技術を用いて、テキストデータベースからオントロジーを自動的に構築する様々な手法が提案されている。その多くは、上位下位関係や部分全体関係など、用語と用語の様々な関係の抽出を目的としたものである。例えば、用語の上位、下位関係を抽出する代表的な手法としては、「A などの B」などの定型表現に着目したものが、「パターン法」と呼ばれている^[1]。この場合、「などの」というパターンの前に出現する名詞句 A を後ろに出現する名詞句 B の下位語として抽出される。しかしながら、幅広い分野の一連の手続きに関する知識をテキストから自動抽出し、それらを体系化する試みはほとんどない。

本研究では、特許から手順オントロジーを自動的に構築する手法を提案する。特許では、新しい技術や発明を説明するために、それを実現する手順を記載することがしばしばある。図 1 は「対訳辞書作成装置」に関する日本国特許（特開 2017-091382）の要約であり、S11 から S16 までの手順から構成されていることがわかる。図 2 は、同じ特許の代表図面であり、要約と同じ内容がフローチャートとして表現されている。

ここで、個々の特許には新規性があるため、ひとつの

特許だけからこれらの情報を抽出しても、それが対訳辞書作成装置の典型的な手順になっているとは限らない。そこで、日本国特許に付与されている分類コードのひとつである F タームに着目し、同一の F タームが付与されている複数の特許から手順情報を抽出し、それらの共通項を検出することで、対訳辞書作成装置の典型的な処理手順に関する知識を自動獲得する。

対訳コーパスから複数の対応文を読み込み S11、複数の対応文から用語を抽出し S12、抽出された用語が用語ペアテーブルに登録されている用語ペアを構成する用語以外である場合には、当該用語を、新規な用語として選定する S13。複数の対応文のマッチングに基づいて、新規な用語のペアを用語ペア候補として取得し S14、用語ペア候補の出現頻度に応じて、当該用語ペア候補を構成する新規な用語ペアを対訳辞書として出力するステップ S16。取得するステップでは、複数の対応文の順序をランダムに変更して前記マッチングを繰り返し行う。

図 1 特許要約における手順の記載例（特開 2017-091382）

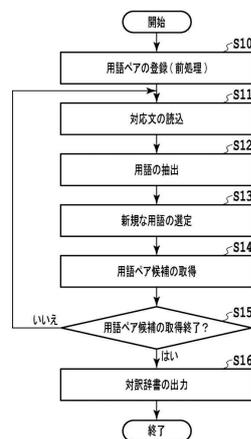


図 2 図 1 の特許要約に対応する代表図面

2 関連研究

2.1 手順情報の抽出

近年、複数の類似した手順テキストから、共通手順を抽出する研究が行われるようになってきている。新森ら^[4]は請求項の構造解析を修辞構造解析の一種と捉え、手掛り語に基づいた請求項構造解析手法を提案している。日本語の請求項には、一般に「～し、～し、～した」のように処理を順序的に記述する順序列挙形式や、「～と、～と、～とからなる」のように、構成要素を列挙する形で記述する構成要素列挙形式など、いくつかの特許固有の記述スタイルが存在する。新森らは、手がかり語と文脈自由文法を用いたルールを使い、日本語の特許請求項の解析を実現している。

これに対し、本研究では、機械学習を導入した要約の構造解析を目指す。近年では、自然言語処理の様々なタスクにおいて深層学習が導入され、その有効性が確認されている。本研究でも、深層学習を用いた要約の構造解析を試みる。

2.2 フローチャート画像の構造解析

フローチャート画像の解析における関連研究プロジェクトとしてCLEF-IPがある^[3]。CLEF(Conference and Labs of the Evaluation Forum)とは、ヨーロッパを中心に行われている情報検索に関するワークショップであり、CLEF-IPは特許を対象としたタスクのことを指す。このタスクは実験レベルだけではなく、現実の課題に即した検索タスクのためのデータセットを提供することで、多言語及びマルチモーダル特許検索タスクの研究の促進を図っている。CLEF-IPでは図形を認識し、フローチャートの要素となるテキスト、エッジ、ノードを検出している。CLEF-IPの基本的な課題は、本研究と共通するが、CLEF-IPが実施された2013年当時と比べ、画像解析技術が大幅に向上している点、また、本研究では、画像データだけでなく、その画像の内容を説明したテキスト(特許要約)も手順情報の抽出に利用する点が異なる。

2.3 画像分類

特許画像に特化し、特許画像の特徴を用いて画像を検索する研究がいくつか提案されている。Kaderら^[2]

は、セマンティック特許画像検索のための新しいマルチモーダルシステムを提案している。画像から視覚的に関連する特徴を抽出するために、深層学習を用いて事前に訓練された埋め込みを使用する。さらに、OCRを用いて画像から数字を抽出し、それをシーンテキストスポットティングに適用して特許テキスト中の言及にマッピングする。次に、CLIPモデルを用いてイラストから構造的な特徴を抽出し、文章変換ツールを用いて、テキスト中の関連箇所からテキスト特徴を抽出する。最後に、テキストと画像の特徴を用いて、表現されたイラストのインデックスを作成する。このように、テキストと画像の両方の特徴を利用することで、特許画像の意味的な類似性を高めることができる。本研究はテキストと画像の両方の特徴を利用するが、画像を検索するためではなく、手順情報を抽出して手順オントロジーを構築する。

3 提案手法

3.1 手順オントロジーの構築手順

本研究では、ひとつの特許から、その代表画像と要約をそれぞれ解析し、手順情報を抽出する。それらを、特許に付与されているFタームごとに集計することで、手順オントロジーを構築する。以下、図3は、機械翻訳(5B091)というテーマコードのFタームの例である。

```

テーマコード：5B091（機械翻訳）
AA
AA00 言語
AA01  ・多言語間
AA11  ・1言語間
AA12  ・・方言・標準語間
...
AB
AB00 処理対象要素
AB01  ・記号、数字、数式
AB11  ・複合語、熟語、イディオム
...

```

図3 Fタームの例(5B091:機械翻訳分野)

この図において、AA00「言語」やAA01「・多言語間」が観点と呼ばれている。観点は階層的な構造をしており、各説明語の前に記述されている「・」が階層の深さを示している。この場合、AA00の下層にAA01やAA11が存在し、さらに、AA11の下にAA12が存在する。

A で始まる観点は、「翻訳の対象」という観点からの分類体系となっている。一方、BA では始まる観点はいずれも「翻訳の方式」に関するものになっている。

図 1 と図 2 には、5B091AA01 という F タームが付与されている。同じ F タームが付与された特許を収集し、それらから手順情報を抽出すれば、多言語間の機械翻訳に関する手順情報の収集が実現できる。本研究では、特許の代表図面と、その図面に関する説明文である要約に着目し、まず、代表図面が手順に関する画像、すなわちフローチャートであるかどうかを認識⁽¹⁾し、次に、各手順の情報を要約⁽²⁾とフローチャート画像⁽³⁾から抽出することで、F タームの観点ごとに手順情報を集約する。3.2 節では (1) フローチャート画像の自動検出について、3.3 節では (2) 要約からの手順の抽出について、3.4 節では (3) フローチャート画像からの手順の抽出について、それぞれ述べる。

3.2 フローチャート画像の自動検出

ある画像がフローチャートであるかどうかを認識するタスクは、画像分類の一種と考えられる。その代表的な手法は、大量のラベル付き画像データを対象に畳み込みニューラルネットワーク (CNN) 等のモデルを用いて学習をする。近年では、このような学習済モデルが公開されており、このモデルを用いてファインチューニングにより目的の画像分類器を構築する手法が一般的になっている。本研究でも、既存のモデルを利用してフローチャート画像の検出器を構築する。

本研究では、ImageNet (<https://www.image-net.org/>) と呼ばれる大規模画像データセットで学習された 7 つの畳み込みニューラルネットワーク (CNN) モデル (VGG16、VGG19、ResNet50、InceptionV3、DenseNet169、DenseNET121、MobileNet) を用いてファインチューニングによる学習モデルを構築し、その有効性について実験により検証する。また、Baseline として、深層学習ライブラリである Keras を用いて Conv2D が 3 層、MaxPooling2D が 2 層の CNN モデルを構築し、ファインチューニングによる手法と比較する。

3.3 要約からの手順情報の抽出

本研究では、要約を入力とし、図 4 に示すような構

造タグ付きの要約を出力することを目的とする。図 4 において、comp タグ、proc タグ、head タグはそれぞれ、構成要素、手順、主題を示す。こうしたタグを自動的に付与するシステムを構築するため、人手で構造タグを付与したデータを準備し、それを教師データとして用いることで、機械学習ベースの構造解析器を構築する。本研究では、要約の構造解析用に新たにタグ付きコーパスを作成するのではなく、請求項の構造解析用タグ付きコーパス^[5]で構造解析器を構築した後、そのシステムを要約に適用する。これは、一般に、要約は、第一請求項と類似した構造を持つことが多いため、請求項の構造解析用のコーパスを本研究に流用できると判断したからである。

```
半導体基板上に、<proc> 半導体膜を形成する工程 </proc> と、前記半導体膜の所定の領域に、<proc> ドーパント不純物を導入する工程 </proc> と、前記 <proc> 半導体膜をパターニングする </proc> ことにより、前記ドーパント不純物が導入された前記半導体膜からなる抵抗素子と、前記ドーパント不純物が導入されていない前記半導体膜からなるゲート電極とを形成する工程とを有することを特徴とする <head> 半導体装置の製造方法 </head>。
```

図 4 請求項へのタグ付与の例

本研究では、近年様々な自然言語処理タスクにおいて、その有効性が確認されている言語モデル BERT を用いて、請求項の構造を解析する。BERT の入力層に請求項を入力し、出力層側で各単語 (トークン) に対応するタグを出力するよう学習する。なお、BERT のモデルは、東北大学が公開している Pretrained Japanese BERT models をそのまま用いた場合と、事前に大量の特許データを用いてファインチューニングをしたモデルの 2 種類で実験を行う。この他、自然言語処理における系列ラベリング問題で広く使われてきた条件付き確率場 (Conditional Random Field: CRF) でも実験を行い、提案手法と比較する。

3.4 フローチャート画像からの手順情報の抽出

フローチャート画像からの手順情報の抽出は、コンピュータビジョン向けライブラリ OpenCV (<https://opencv.org>) を用い、1. 輪郭の抽出、2. 輪郭の近似・

座標情報の抽出、3. 形状の認識、4. 面積によるクラスタリング、5. 文字認識、6. フローチャートと本文の対応付けの6つの手順から構成される。

図5では、形状認識の際に形状ごとに色分けをしており、長方形は緑、楕円形は青、ひし形は赤で表している。解析結果からわかるとおり、矩形領域、条件分岐を示すひし形、フローの最初と最後を示す楕円が正しく解析されている。一方で、条件分岐が連続する個所の右側の領域など、抽出すべきでない個所を誤って抽出してしまっている。

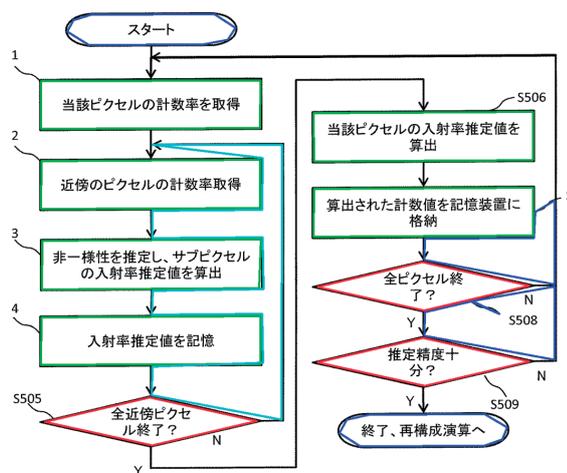


図5 手順情報の抽出結果

4 実験

提案手法の有効性を確認するため、フローチャート画像の自動検出、請求項の構造解析、フローチャート画像の解析に関する実験を行った。それぞれ、4.1節、4.2節、4.3節で報告する。

4.1 フローチャート画像の自動検出

実験データ

2018年の公開公報から抽出した7,099画像に対し、人手で(1)フローチャート、(2)表、(3)構成図、(4)その他に分類したデータを用いる。

実験手法

3.2節で述べたとおり、ImageNetで学習された7つのCNNモデルを用いて、ファインチューニングによるフローチャート画像の検出器を構築した。Baselineとして、Conv2Dが3層、MaxPooling2Dが2層の畳み込みニューラルネットワーク学習モデルを構築し

た。評価は精度、再現率、F-measureを用いた。

実験結果

実験結果を表1に示す。表より、今回構築した学習モデルの中では、精度ではDenseNet121が最もフローチャートの検出精度が高いことが分かった。

表1 8つのモデルによるフローチャートの検出精度

	精度	再現率	F-measure
Baseline	0.8508	0.8902	0.8701
VGG16	0.8750	0.9711	0.9205
VGG19	0.9227	0.9653	0.9435
ResNet50	0.8698	0.9653	0.9151
InceptionV3	0.9422	0.9422	0.9422
MobileNet	0.9326	0.9595	0.9459
DenseNet169	0.9593	0.9538	0.9565
DenseNet121	0.9645	0.9422	0.9532

4.2 要約からの手順情報の抽出

実験データ

日本国特許の請求項2456件に対し、人手でhead, proc, compタグを付与し、さらにブロック間の依存関係を付与したデータを用いる。

実験方法

人手によるタグ付きデータのうち、3/4を訓練用とし、残りの1/4を評価用に用いた。評価には、精度、再現率、F-measureを用いた。

比較手法

以下の3種類の手法で実験を行った。

- BERT (事前学習なし) : Pretrained Japanese BERT models をそのまま利用
- BERT (事前学習あり) : 公開特許公報から任意に選択した350万文を用いてMLMタスクにより事前学習したモデルを利用
- CRF (ベースライン手法) : 前後4単語のユニグラム、バイグラム、トライグラムを素性として利用

実験結果

結果を表2に示す。表より、CRFと比べ、BERT(事前学習なし)が再現率を0.09以上向上させることができた。一方、BERT(事前学習あり)は、BERT(事前学習あり)の精度を若干向上させることができたものの、F-measureではBERT(事前学習なし)とほぼ同値となった。特許を用いた事前学習がそれほど解析精度の向

上に大きく貢献しなかった理由のひとつは、事前学習に用いた文の多くは請求項ではなく明細から抽出してきたからであろうと思われる。このため、文体が異なる請求項に対して、大きな改善が見込めなかったためと考えられる。

表2 請求項の構造解析精度

手法	精度	再現率	F-measure
BERT(事前学習なし)	0.773	0.867	0.817
BERT(事前学習あり)	0.791	0.837	0.814
CRF(ベースライン手法)	0.816	0.776	0.795

4.3 フローチャート画像からの手順情報の抽出

実験データ

2018年の公開公報から抽出した37画像(いずれもフローチャート)を用いる。これらの画像に対し、人手で502ノードの座標を判定した。なお、各座標はオブジェクトの形状に関係なく、すべて矩形領域の座標として近似的に判定している。

評価方法

出力結果と比較し性能を評価する。ノードの座標の一致度の評価尺度として精度と再現率を用いる。なお、システムの出力と人手のノードを比較する際、領域の重なる面積が90%を超えた時、2つのノードが一致したと判定する。また、比較手法としてクラスタリングをしない場合の精度、再現率を求める。

比較手法

3.1節で述べた6つの手順のうち、手順4においてクラスタリングを行った場合(提案手法)とクラスタリングを行わず、すべてのノードを抽出した場合(ベースライン手法)とで比較した。

実験結果

結果を表3に示す。表より、提案手法と比較手法を比べると再現率を低下させることなく高い精度を得ることができたため、クラスタリングは有効であり、提案手法の方が優れているという結果が得られた。

表3 フローチャート画像からの手順情報の抽出結果

	精度	再現率
(1) 提案手法	0.753	0.673
(2) ベースライン手法	0.693	0.673

5 おわりに

本研究では、日本国特許の要約および代表図面から手順情報を抽出し、さらに、特許に付与されているFタームと組み合わせることで、手順オントロジーを構築する手法を提案した。実験の結果、フローチャート画像の検出では、DenseNet121を用いてファインチューニングしたモデルで精度0.9645を達成した。請求項の構造解析ではBERT(事前学習なし)手法によりF-measureで0.817が得られた。また、フローチャート画像からの手順情報の抽出では、精度0.753、再現率0.673を達成した。

参考文献

- [1] Hearst, M. A., Automatic Acquisition of Hyponyms from Large Text Corpora, in Proceedings of the 14th International Conference on Computational Linguistics, pp. 539-545, 1992.
- [2] Kader, P. et al., A Multimodal Approach for Semantic Patent Image Retrieval. Patent Text Mining and Semantic Technologies 2021, 2021.
- [3] Piroi, F. et al., Overview of CLEF-IP 2013 Lab Information Retrieval in the Patent Domain, Information Access Evaluation. Multilinguality, Multimodality, and Visualization. CLEF 2013. Lecture Notes in Computer Science, Vol. 8138. Springer, Berlin, Heidelberg, 2013.
- [4] 新森昭宏他, 手がかり句を用いた特許請求項の構造解析, 情報処理学会論文誌, Vol. 45, No. 3, pp. 891-905, 2004.
- [5] 難波英嗣, 手順オントロジー構築のための特許請求項の構造解析, 情報処理学会第138回IFAT研究会, 2020.

