

令和 2 年 5 月 28 日現在

機関番号：14501  
研究種目：基盤研究(C) (一般)  
研究期間：2017～2019  
課題番号：17K00236  
研究課題名(和文) 知識グラフを組み込んだニューラルネットワークによる物体認識 物体・意味写像の提案

研究課題名(英文) Object Recognition by Deep Neural Network with Knowledge Graph Embedding -  
Proposal for Semantic Object Projection -

研究代表者  
有木 康雄 (ARIKI, Yasuo)  
神戸大学・都市安全研究センター・名誉教授

研究者番号：10135519  
交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：本研究では、物体の画像特徴と物体に関する知識を同時に学習することにより、既知物体と未知物体の認識精度向上を目指すゼロショット学習について研究を行った。この結果、知識グラフは単語やテキストといった記述レベルより優れており、完全なWordnet知識グラフを用いるモデルでは、現在最も優れている方法より45%も精度を改善することができた。また、深層学習を用いたゼロショット学習の応用として、知識を付加した対話システムについて研究を行った。

#### 研究成果の学術的意義や社会的意義

物体画像だけから特徴を抽出し認識するという方法には限界があり、「物体に関する知識も同時に学習させることにより、物体認識精度の向上、未知物体の識別と学習(ゼロショット学習)が可能になる」という考えに基づき、シンボル表現された知識と画像信号とを融合して学習を進める方法を明らかにした。また、この考えに基づき、発話文と知識を融合することで、知的な対話を行う方法を明らかにした。

研究成果の概要(英文)：In this study, zero-shot deep learning was investigated, which can learn both of image features and the related knowledges simultaneously, and then improve the recognition accuracy of known objects as well as unknown objects. As the results, knowledge graph was proven to be superior to words and texts in terms of the description ability and the knowledge graph of the complete Wordnet improved the recognition rate by 45% compared with state of the arts. In addition, as the application of zero-shot deep learning, we studied about dialogue systems which can incorporate external knowledge.

研究分野：知能情報学

キーワード：ゼロショット学習 意味的特徴 画像特徴 知識グラフ 深層学習 意味ベクトル 対話システム

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

(1) 近年、Deep Neural Network(DNN)の出現により、物体探索、認識、検索技術は大幅に精度が向上し、大量の画像データベースImageNetを使った物体認識コンテストが開催され、年々その精度が向上している。しかし、その認識方法は画像特徴だけに基づいて、よく似た物体の名称を出力しているに過ぎず、このような方法では認識精度に限界が生じる。

(2) この問題に対して、物体の画像だけでなく物体に関する知識も含めて学習することにより、認識精度が向上すると考えられる。例えば、「猫」の学習では、「猫」の画像を大量に集めるだけでなく、「猫」に関する知識、「猫は地球上の生物である」、「猫は哺乳動物である」、「哺乳動物は脊椎動物である」、「哺乳動物は毛におおわれている」などの知識も同時に学習させる。これらの知識はセマンティックなグラフ(知識グラフ)として表現でき、性質や構造に関する多くの情報を表現している。このように、物体に関する知識も同時に学習させることにより、知識に対応する物体特徴や構造を明示的に抽出できるため、認識精度の向上が期待できる。

(3) 物体画像だけから特徴を抽出すると、未知の物体は特徴がよく似た既知物体のどれかに識別されるため、未知物体として識別することは難しい。しかし、各物体にはどのような構造があり、どのような性質を持っているのかという知識(属性)も同時に学習しておくことにより、既知物体の識別精度が向上するだけでなく、未知物体と既知物体の差も明確になり、未知物体の識別が可能になると考えられる。いわゆるゼロショット学習(一度も学習していないものを識別すること)が可能になる。

### 2. 研究の目的

(1) 近年、深層学習の発展により、物体認識の精度が大きく向上している。しかし、その方法は画像特徴のみに基づいているため、認識精度に限界があり、未知物体の識別も困難であるといった問題がある。これに対して、本研究では、物体の画像特徴だけでなく、物体に関する知識も同時に学習させることにより、既知物体の認識精度と未知物体の識別率の向上を目的としている。

(2) 物体の画像特徴だけでなく、物体に関する知識も同時に学習させ、既知物体の認識精度と未知物体の識別率を向上させる研究は、一般的にゼロショット学習と呼ばれている。深層学習を用いたゼロショット学習は応用範囲が広く、画像だけでなく、言語処理に関しても拡張可能である。本研究では、画像のゼロショット学習と同様、言語におけるゼロショット学習の研究を進めることで、人とロボットの円滑なコミュニケーションに関する研究を展開する。

### 3. 研究の方法

(1) 物体画像の学習データセットを決定し、物体画像から画像特徴ベクトルをCNNにより学習する。また、物体の知識グラフを決定し、知識グラフから属性特徴ベクトルをDNNにより学習する。その後、画像特徴ベクトルから属性特徴ベクトルへの写像を、DNNにより学習する。この写像を用いて、ゼロショット学習が可能かどうかについて、検証する。

(2) 深層学習を用いたゼロショット学習の応用として、下記内容の研究を進める

画像のゼロショットセグメンテーション

知識と対応付けて人と論議できる対話システム

発話意図に関係する知識を検索して対話するシステム

大規模知識グラフでユーザ入力文を拡張して対話するシステム

意味ベクトルを付加した発話スタイルに頑健な対話システム

### 4. 研究成果

(1) 意味表現がゼロショット学習の精度に及ぼす影響の解析

異なる意味表現によって、ゼロショット学習の精度がどのように変化するか、最も精度の高い意味表現はどのようなものであるかについて研究を行った。まず、一般物体のゼロショット学習に対して、標準的な評価ベンチマークとなっているImageNetをベンチマークとして用いた。ImageNetの各クラスはWordnetの概念と対応しており、3つの要素で定義されている。それらは、見出し単語、見出し単語を自然言語で説明したテキスト文書、Wordnetの他ノードに述語を介してリンクしている知識グラフである。同じ意味表現でも、コンピュータで扱えるベクトル形式に変換する方法が種々提案されているので、異なる意味表現、異なる変換方法によって、ゼロショット学習の精度がどのような影響を受けるか調べた。その結果、次のような結果が得られた。

単語表現は変換方法によって精度が同じではないことを発見した。従来のWord2vecの代わりにGloVeを用いると、ベースラインモデルの精度を2倍にすることができた。

知識グラフは他の記述レベルより優れている。完全なwordnet知識グラフを用いるモデルでは、現在最も優れている方法より45%も精度を改善することができた。

現在、最も研究が行われているテキスト文書は、ゼロショット学習のベンチマークでは、良い結果が得られなかった。

#### (2) 新たなベンチマークの提案

これまでのゼロショット学習の精度が、予想外に低い理由について考察した。その結果、ゼロショット学習のベンチマークとして用いていたImageNetのデータベースに、ゼロショット学習にとって致命的な欠陥があり、これを改善することで、ゼロショット学習の精度が大幅に改善されることが分かった。まず、この欠陥をなくすために、ゼロショット学習データベースに構造的バイアスを導入した。また、この構造的バイアスを導入した新たなベンチマークテストのためのデータベースを自動構築する方法について提案した。

#### (3) 画像セグメンテーションへの応用

ゼロショット学習による物体認識は、入力対象の画像特徴から対象の意味的属性特徴への写像を学習するものである。この写像を画素ごとに求めると、画像中に存在する物体を、画素ごとに分割することができる。1枚の画像に映っている未知物体の名称を出力するゼロショット学習に対して、1枚の画像に映っている未知物体を画素ごとにセグメンテーションするゼロショットセグメンテーションを提案した。

#### (4) 議論対話への応用

あるトピックに関して、人と知的に対話できるロボットについて研究を行った。ゼロショット学習は、画像特徴と意味的属性特徴との対応付けにより、未知物体を認識する問題である。同様に、人の意見に含まれる言語特徴と、その周辺に存在する知識や概念の意味的属性特徴を対応付けることによって、人の意見とは逆の意見を、深層学習によって生成することができる。こうして、ゼロショット学習の拡張として、人の意見に従うロボットとの対話ではなく、人と対立する意見を持ち、論議できる対話ロボットについて研究した。

#### (5) 人とロボットとの知的対話への応用

人と円滑なコミュニケーションを行うためには、人の発話を言語特徴として聞き取るだけでなく、その発話周辺の知識や概念の意味的属性特徴と対応付け、新たな内容を展開していく必要がある。言語特徴と未知なる知識の意味的属性とを対応付けるため、ゼロショット学習の一種と考えられる。人の発話の言語特徴として、LSTM Encoder-Decoder を用いて、ユーザの発話からその発話意図とキーワードを推定した。また、知識ベース中の文書を知識の意味的属性と考え、発話意図とキーワードに関係する知識を検索し、その結果をLSTM Encoder-Decoder を用いて、システムの発話を生成する研究を行った。

#### (6) 単語概念の拡張に基づく雑談対話

LSTM Encoder-Decoderによる雑談対話システムの汎化性能を向上させる研究を行った。雑談は幅広い話題を扱い、表現も多種多様であるため、汎化性能を向上させることが困難である。ユーザが、あまり現れないマイナーな単語や表現を入力した場合、意味的に関係性の低い応答を生成する可能性がある。この問題に対して、ゼロショット学習と同様に、ユーザ入力文に含まれる単語の分散表現に、大規模知識グラフである日本語WordNetの上位語の分散表現を加算する。こうして、学習が不足している単語や未知語が入力されても、単語を概念的に拡張して、適切な応答を生成する方法を提案した。

#### (7) 意味ベクトルの抽出に基づく対話

Transformer Encoder-Decoderを用いた翻訳や対話システムにおいて、発話スタイルに頑健な方法を研究した。システムの学習に用いたデータと実際のユーザ発話の間には、「標準語と方言」、「話し言葉と書き言葉」、「敬体と常体」などといった発話スタイルのギャップが存在する。このため、Transformer Encoder-Decoderは適切な意味表現が生成できなくなり性能が低下する。この問題に対して、学習データとは異なるデータベース・知識ベースを併用して、敵対的生成ネットワークにより、発話スタイルに依存しない意味ベクトルと、発話スタイルベクトルに分解する手法を提案した。この意味ベクトルを付加することにより、発話スタイルに頑健な翻訳・対話システムを構築した。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1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 Tristan Hascoet, Yasuo Ariki and Tetsuya Takiguchi	4. 巻 -
2. 論文標題 Semantic embeddings of generic objects for zero-shot learning	5. 発行年 2019年
3. 雑誌名 EURASIP Journal on Image and Video Processing	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） doi.org/10.1186/s13640-018-0371-x	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32件（うち招待講演 0件 / うち国際学会 13件）

1. 発表者名 Tristan Hascoet, Yasuo Ariki, Tetsuya Takiguchi
2. 発表標題 On Zero-Shot Recognition of Genetic Objects
3. 学会等名 International Conference on Computer Vision and Pattern Recognition (CVPR) (国際学会)
4. 発表年 2019年

1. 発表者名 Kazuaki Furumai, Tetsuya Takiguchi, Yasuo Ariki
2. 発表標題 Generation of Objections Using Topic and Claim Information in Debate Dialogue System
3. 学会等名 International Workshop on Spoken Dialog System Technology (IWSDS) (国際学会)
4. 発表年 2019年

1. 発表者名 古舞千暁, 滝口哲也, 有木康雄
2. 発表標題 発話スタイルに非依存な文章表現を用いたスタイルロバスタなSeq2seqモデルの提案
3. 学会等名 言語処理学会第26回年次大会発表論文集
4. 発表年 2020年

1. 発表者名 麻生大聖, 高島遼一, 滝口哲也, 有木康雄
2. 発表標題 WordNetを用いた雑談対話システムの汎化性能の向上
3. 学会等名 電子情報通信学会技術研究報告
4. 発表年 2019年

1. 発表者名 Yuki Matsuyoshi, Tetsuya Takiguchi, Yasuo Arika
2. 発表標題 User 's Intention Understanding in Question-Answering System Using Attention-based LSTM
3. 学会等名 APSIPA (国際学会)
4. 発表年 2018年

1. 発表者名 谷田 啓一, Tristan Hascoet, 滝口 哲也, 有木 康雄
2. 発表標題 ゼロショット学習を用いた一般物体セグメンテーション
3. 学会等名 情報処理学会第81回全国大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

神戸大学 システム情報学研究所 情報科学専攻 メディア情報講座  
<http://www.me.cs.scitec.kobe-u.ac.jp/~takigu/publications.html>

## 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
研究協力者	ハスコエツ トリスタン  (HASCOET Tristan)		
研究協力者	古舞 千暁  (FURUMAI Kazuaki)		
研究協力者	谷田 啓一  (TANIDA Keiichi)		
研究協力者	麻生 大聖  (ASO Taisei)		
研究協力者	松好 祐紀  (MATSUYOSHI Yuki)		