

IoTデータ向け マルチモーダル深層学習基盤

IoTデータ向けマルチモーダル深層学習基盤は、画像とテキストデータや数値データを組み合わせてAIによる学習を行うことにより、高精度な類似画像検索や画像分類、オブジェクト検出などを実現する学習基盤である。適用技術として、Tripletネットワーク学習および言語情報の分散表現を活用した画像データを対象とする深層学習技術を紹介する。また、ソリューション展開に向けて想定される適用分野や顧客への提供価値について述べる。

池浦 康充 | Ikeura Yasumichi

岡本 光一 | Okamoto Koichi

加嶋 亮平 | Kashima Ryohei

土方 悠介 | Hijikata Yusuke

廣池 敦 | Hiroike Atsushi

1. はじめに

各種記憶装置の容量の大規模化、計算機演算性能の向上、ネットワーク上のデータ転送速度の高速化などにより、各種業務において、大量の電子的履歴情報の蓄積が可能となった。情報の内容としても、従来の書誌的情報だけではなく、映像、画像などのメディア情報、各種センサー情報 [IoT (Internet of Things) データ] などの多様な電子情報も蓄積することが可能となっている。このような背景から、これらの電子的履歴情報を利活用するための情報処理技術へのニーズが高まっている。例えば、特許、商標、意匠などの審査業務では、出願された文献、関連する図面、審査業務に関連し付与された書誌的情報などの多様な情報が業務の履歴情報として蓄積さ

れる。これらの履歴情報の活用を実現し、審査業務の効率化、出願に関わる作業の効率化が実現できれば、産業分野の発展に大きく寄与することとなる。

一方、メディア情報の情報処理技術として近年脚光を浴びているのが深層型ニューラルネットワークである。特に、画像・映像認識の分野では、従来、解決が困難と考えられるような課題を解決するモデルが次々に提案されている。ただし、学術分野で成果を上げたモデルを、そのまま実業務システムに適用すると、さまざまな困難に直面することが多い。例えば、実業務における履歴情報は、学術分野でのベンチマーク用データとは異なり、多様性、例外、不完全性などにより、そのままモデルを適用することが困難な場合が多い。そこで日立では、現在、実業務に対して、深層型ニューラルネットワークの学習技術を適用するための基盤技術（フレームワーク）および学習モデルの研究開発を推進している。

2. IoTデータ向け マルチモーダル学習基盤の概要

本章では、マルチモーダル深層学習基盤の概要について説明する。図1にマルチモーダル深層学習基盤の概要を示す。マルチモーダル深層学習基盤は、深層学習技術の活用により「類似画像検索」、「画像分類」、「オブジェクト検出」などの画像認識に係るサービスを有している。これらの機能を活用することにより、従来、人手により大量に画像を取り扱い、画像の検索や分類を行っていた業務について、画像の検索や分類を自動化することで、業務の効率化を図ることが可能となっている。また、深層学習の機能に関しても、学習の実行や学習進捗状況についてGUI（Graphical User Interface）ベースで操作・確認することが可能となっており、機械学習に関して未習熟であっても、容易に操作・確認することができ、機械学習に関する画像認識技術のPoC（Proof of Concept）をスムーズに実施することが可能な環境を提供する。

実際の学習における運用では、学習の実行だけではなく、学習データを作成する（いわゆる、データに対して「メタ付け」を行う）アプリケーションとの連携、学習結果の実ソリューションへの反映などのアプリケーション間の連携が必要となる。本学習基盤では、「EnraEnra」^{1), 2)}のアプリケーションフレームワークを活用することによって、「(a)メタ付け」、「(b)学習」、「(c)ソリューション展開」という一連の処理を効率よく実現する、統一的な

アプリケーションを提供する。以下にその主な特徴を示す。

(1) 「EnraEnra」を活用した類似画像検索

日立が開発した類似画像検索システムである「EnraEnra」は、画像特徴量の抽出技術などの画像認識技術およびデータベース技術、プラットフォーム技術の三つの分野の技術から構成される。この中で、データベース技術とプラットフォーム技術は、深層型ニューラルネットワークと相性が良い。

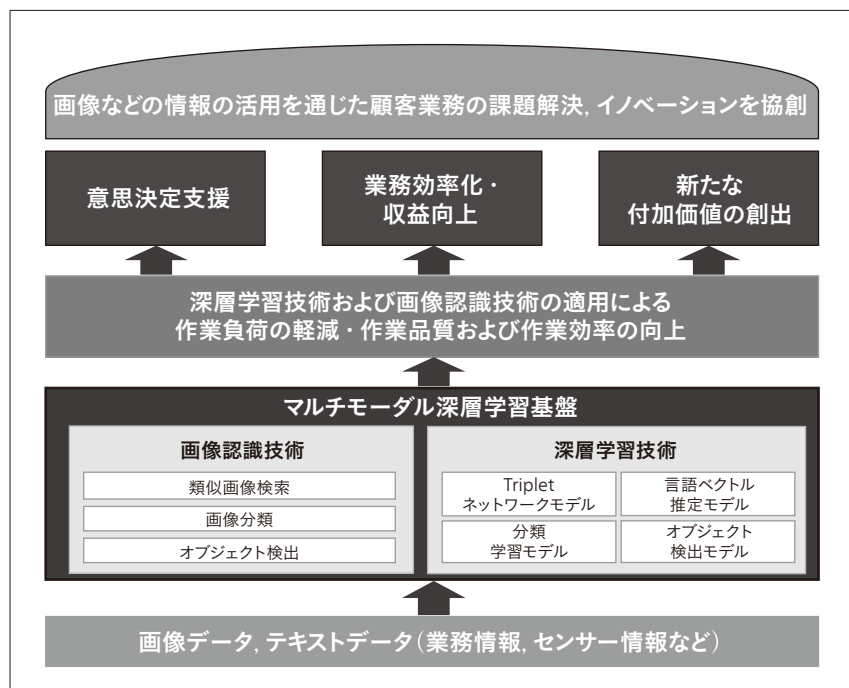
深層型ニューラルネットワークは、実は「ベクトルからベクトルを生成する仕掛け」と見なすことができる。「EnraEnra」の最大の特長は、高速の類似ベクトル検索機能である。「EnraEnra」では、独自のクラスタリングアルゴリズムに基づく「近似近傍ベクトル探索」をDBMS（Database Management System）の標準機能として提供しており、数十から数千次元程度のベクトルで構成される大量のデータ集合から、類似したデータを高速で検索することができる。深層型ニューラルネットワークは、まさに、このようなベクトルを出力するもので、類似検索と組み合わせることによって、多様な機能が提供可能となる。

(2) さまざまな学習モデルに対応

マルチモーダル深層学習基盤は複数の学習モデルを備えており、上記のような類似画像検索に使用する画像の特徴量を抽出することが可能なモデルの他に、画像分類用の学習モデルやオブジェクト検出用の学習モデルも備えている。このため、画像分類やオブジェクト検出を行

図1| マルチモーダル深層学習基盤の概要図

顧客が保有する業務情報やセンサー情報などを基に深層学習を実施することで、より高精度な類似画像検索、画像分類、オブジェクト検出が可能となる。



うことが可能となっている。この他にも、独自性のある Triplet ネットワークモデルやマルチモーダル学習を行う可変個数言語ベクトル推定モデルも備えている。モデルの詳細については次章で記述する。

3. 学習モデルの実装例

本章では、マルチモーダル学習基盤が保有する特徴的な学習モデルについて説明する。

3.1

Triplet ネットワークモデル

Triplet ネットワークモデルとは、あるクエリに対して、検索結果に含まれてほしいデータを正事例、含まれてほしくないデータを負事例としたうえで、クエリ、正事例、負事例の3者の組み合わせ (triplet) のベクトル空間上での距離関係に基づいた損失関数が小さくなるように特徴量ベクトルを学習するモデルである。

ある学習用のデータセットが与えられたとき、この triplet は膨大な数となる。Triplet ネットワークモデルの学習では、triplet をいかに選定するかが重要な課題と言える。本学習基盤では、類似ベクトル検索を使うことによって学習に有効な triplet を自動的に選定する学習方式を実現している。図2に、1エポック (学習時に学習セットを一巡する1周期) 内での学習処理の概略を示した。

最初に、その時点のネットワークの状態を用いて全画像の特徴量ベクトルの抽出を行う (「画像特徴量ベクトル抽出」)。この特徴量ベクトルをクラスタリングしたうえで、全画像を対象とした類似検索を、全画像中から順番にクエリを選択し、実行していく。各画像の類似検索結果は、一般に上位に正事例が集まり、負事例は検索され

ないか、下位に出現した方が望ましい。そこで、正事例より上位に出現した負事例を選択するように triplet を構成していく。このようにして構成された triplet を用いて学習を行うと、検索結果中の正事例はより上位に、負事例はより下位に移動していくように状態が変わっていく。

監視映像中の人物検索に関する評価実験で、この学習方式が分類学習と比較して、より高い精度の特徴量ベクトルを構成できることを確認している。

3.2

可変個数言語ベクトル推定モデル

近年、言語情報の表現として、word2vec^{※)}による単語の分散表現 (ベクトル表現) が一般的に利用されるようになった。ここでは、画像から word2vec が出力する単語ベクトルを推定するモデルについて紹介する (図3参照)。

このモデルの適用は、画像に対して、任意個数の検索用キーワードが付与されたような状況が想定される。例えば、単語ベクトルの次元数を d とすれば、ニューラルネットワークの方は、画像を入力すると、 d 次元のベクトルを n 個出力するように設計する。一方、キーワードの方は、単語とベクトルの対応付け関係をデータベース化しておくことにより、 d 次元の単語ベクトルの集合に変換される。キーワードの個数を m 個とし、次に、学習過程の各時点での出力ベクトルと単語ベクトルのすべての組み合わせについてベクトル間距離を算出する。この算出された距離が小さい順に対応付けを行い、対応付けされた組の距離の総和が小さくなるように学習を行う。なお、多くの場合、 $m < n$ となるため、単語ベクトルとの対応付けができなかった出力ベクトルが残ってしま

※) 大量のテキストデータを解析し、各単語の意味や文法を捉えるために単語を低次元 (数百次元) の数値ベクトルへ変換表現する手法。ベクトル表現から単語同士の足し算・引き算による類推や類似度計算を行うことができるため、単語のラベル推定や分類などのタスクにて使用されている。

図2| Triplet ネットワークの学習

各エポックのネットワークの状態の特徴量抽出と類似ベクトル検索を実行することによって、学習に最適な triplet を選択する。

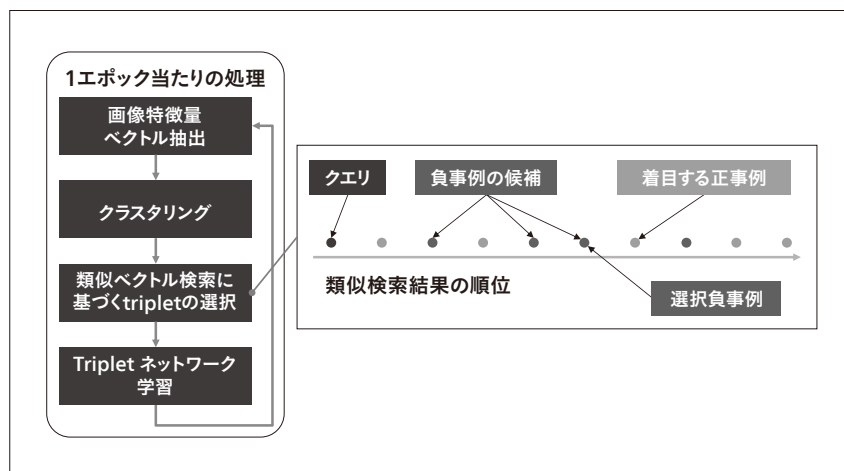
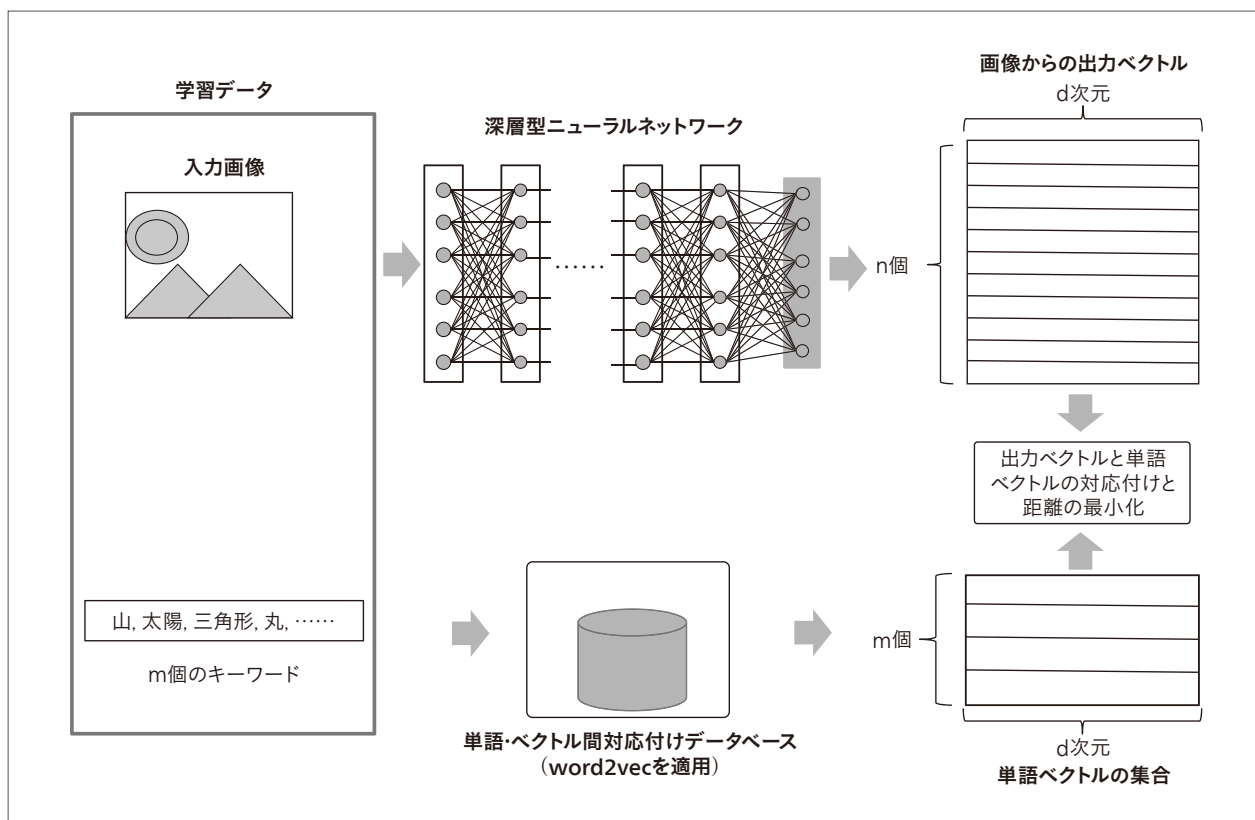


図3| 可変個数言語ベクトル推定モデル

画像に付与されたキーワードをword2vecによりベクトルに変換したうえで、深層型ニューラルネットワークによる推定を行う。一定の個数である深層型ニューラルネットワーク側の出力ベクトルと、可変個数のキーワードとの対応付けは、その時点の推定結果に応じて動的に選択される。



う。このようなベクトルについては、0ベクトルに近づくように学習する。

このモデルで実際に推定を行う際には、単語ベクトルのデータベースに対して、モデルの出力ベクトルをクエリとする類似検索を行う。検索上位にきたものが、入力された画像に付与すべき単語となる。

4. 適用事例

本章においては、マルチモーダル深層学習基盤を活用し、業務効率化を実現することが想定される適用事例を紹介する。

4.1 先行技術調査業務への適用

現在の特許の先行技術調査業務においては、文献中のテキスト情報や文献に付与された分類情報を用いて検索を行い、調査対象の文献の絞り込みを行っている。絞り込んだ文献に対して人手によりスクリーニングを実施するため、調査対象文献が少ないほど、効率的に業務を行うことが可能である。

日立では、2016年より特許庁のAI(Artificial Intelligence)技術の活用による先行技術調査業務の効率化に関する調査研究事業を行っており、特許図面の分類、検索についても研究を行ってきた。

本深層学習基盤の類似画像検索を活用し、図面の情報を用いた直接的な検索を行うことにより、人手によるスクリーニングが不要な文献を調査対象から除去することが可能となり、業務の効率化を図ることができると考えられる。

4.2 橋梁などの点検・診断業務支援への適用

橋梁などの点検・診断業務については、過去にひび割れなどの損傷パターンに対して人手で診断を行った大量の画像データが蓄積されている。この点検・診断業務に本深層学習基盤を適用し、これまで蓄積された診断済みの画像データと診断結果のテキスト情報を学習用データとして用いて、画像分類により、改修が必要なひび割れを自動的に分類し、抜け漏れなく損傷の種類、程度を確実に記録するための支援を行う。これにより、状態の良い橋梁は点検・診断の人的リソースをかけ過ぎることなく確実に対処することで、重要度・緊急度の高い橋梁に

人的リソースを注力できるようになる。

4.3

高速人物発見・追跡ソリューションとの連携

高速人物発見・追跡ソリューションはAI技術を使った画像解析により、人物の特徴や全身画像を使って特定人物の発見・追跡を支援するもので、警察での捜査業務などへの適用を検討している³⁾。高速人物発見・追跡ソリューションとの連携は2021年度から計画している。当該ソリューションと連携し、学習結果を応用することで、より顧客の環境に合わせた人物の発見・追跡が可能となることが期待される。

5. おわりに

本稿では、現在、日立が開発を行っているマルチモーダル深層学習基盤と、適用されている技術や適用分野について紹介した。

今後も、PoCなどを通じて得た成果を基に、顧客業務の効率化に通じる新たな機能追加を実施し、顧客に対して高付加価値なサービスを適用していく所存である。

参考文献

- 1) 廣池敦，外：大規模に蓄積された画像および音声を対象とする情報検索，日立評論，95，2，200～205（2013.2）
- 2) 廣池敦：類似画像検索システム「EnraEnra」，人工知能学会誌，29巻，5号，pp.430～438（2014.9）
- 3) 沖田英樹，外：安全・安心な社会の構築に貢献するAI技術を用いた映像解析ソリューション，日立評論，102，3，400～406（2020.7）

執筆者紹介



池浦 康充

日立製作所 社会ビジネスユニット 公共システム事業部
官公システム第三部 所属
現在，マルチモーダル深層学習基盤の開発に従事
プロジェクトマネジメント学会会員



岡本 光一

日立製作所 社会ビジネスユニット 公共システム事業部
官公システム第三部 所属
現在，マルチモーダル深層学習基盤の開発に従事



加嶋 亮平

日立製作所 社会ビジネスユニット 公共システム事業部
官公システム第三部 所属
現在，マルチモーダル深層学習基盤の開発に従事



土方 悠介

日立製作所 社会ビジネスユニット 公共システム事業部
官公システム第三部 所属
現在，マルチモーダル深層学習基盤の開発に従事



廣池 敦

日立製作所 研究開発グループ
データサイエンスラボラトリ 映像解析ソリューションPJ 所属
現在，AI技術を活用した画像検索システムの研究開発に従事
博士（工学）
日本心理学会会員