



2021年度 グループ 3 活動報告

研究開発項目 3 : 人間の知識・概念獲得の研究開発

研究開発課題 1 : 概念理解とマルチモーダル認識の研究開発

(1) 未知・新規な状況を判断する技術開発

未知・新規な状況を判断するためには、収集されたデータに基づいて、生成モデルを作成することが重要となる。生成モデルを構築できれば、新規入力データとモデルの差分を計算することで未知・新規かどうかを判断可能となる。そこで、本年度は CA への搭載も見越した超効率的な生成モデルを構築した。我々はスパイクニューラルネットワーク (SNN) に着目した。SNN はニューロモルフィックデバイス上で超高速かつ超低消費電力で動作させることが可能である。そのため、SNN は CA 上で動作する生成モデルとして、様々な応用が期待される。本研究では、画像生成を可能とするために、SNN を用いた変分オートエンコーダ (VAE) を構築した。VAE は生成モデルの中でも安定性が高いことで知られており、近年はその品質が向上している。通常の VAE では、潜在空間は正規分布で表現され、サンプリングの際に浮動小数点演算が必要となる。しかし、SNN ではすべての特徴量が 2 値の時系列データである必要があるため、これは不可能である。そこで、自己回帰型 SNN モデルで潜在空間を構築し、その出力からランダムにサンプルを選択して潜在変数をサンプリングすることで、潜在変数がベルヌーイ過程に従うようになり、変分学習が可能となる。SNN 層のみで VAE を構築したのは本研究が最初である。この研究は、研究開発課題 4 (棕田) と共同で実施した。

実験では、手書き数字データセット MNIST とファッション画像データセット Fashion MNIST、物体画像データセット CIFAR10、さらに顔画像データセット CelebA を使用した。いずれのデータセットにおいても、再構成損失やインセプションスコアにおいて、従来のニューラルネットワークと比較し、高い性能を得られることが分かった。本モデルをニューロモルフィックデバイス上で動作させることで、速度を 100 倍程度高速化でき、消費電力を 100,000 分の 1 程度に削減可能だけでなく、再構成性能においても従来のネットワークを超えるものができており、マイルストーンを大きく超えた想定外の成果であるといえる。

(2) 未知物体や状況に関する情報を獲得する技術開発

未知物体や状況に関する情報を獲得する技術の研究と開発を開始した。本技術課題を進めるには未知な情報を言語化し質問を生成すること、質問の意図を明確に回答者に伝えること、冗長な回答の中から欲しい情報のみを適切に抽出することが重要となるために、それぞれの項目に対して、課題推進者が開発してきたアルゴリズムの問題点を見直した手法の開発に着手した。

今年度は特に視覚的質問応答 (Visual Question Answering, VQA) に着目した。従来の VQA モデルは、世界の知識に対する推論を必要とする問題では、誤った推論をする傾向がある。最近の研究では、推論問題とともに低レベルの知覚情報を提供する質問で VQA モデルを訓練すると性能が向上することが示されている。これに着想を得て、本年度は、正しい推論に有用な補助的な知覚情報を積極的に獲得するための質問を生成する新しい VQA モデルを考案した。

本 VQA モデルにおいて、訓練データに含まれる回答候補数という観点からは 5000 クラスであり、VQA モデルが回答可能なクラス数ということであれば、約 1000 クラスとなっており、当該年度のマイルストーンを十分にクリアしている。

(3) 少数の教師情報からの知識や概念を構築する技術開発

少数の教師情報からの知識や概念を構築する技術の研究と開発を開始した。特に、本年度は研究項目 (1) と (2) に関連の深い、視覚情報に関する質問から得られる回答を用いて、視覚情報の質問応答の知識を拡張する手法やシステムの構築に着手した。人間から得られる回答量はその負担を考慮すると多くはできないため、少数の回答つまり少量の教師情報

から知識や概念を効率的に構築する手法が求められる。また、逐次的に得られた回答から知識の構造を構築し再利用する必要がある。

そこで本年度は、視覚的質問生成（Visual Question Generation, VQG）に着目した。VQG は、画像から質問を生成するタスクである。人間が画像に対して質問をするとき、その目的は多くの場合、何らかの新しい知識を得ることである。しかし、VQG に関する既存の研究は、主に答えや質問カテゴリからの質問生成を扱っており、知識の獲得という目的を見落としている。そこで、我々は知識獲得の視点を VQG に導入するために、K-VQG と呼ばれる新しい知識獲得型 VQG データセットの構築を実施した。

構築した K-VQG データセットは、既存の知識認識型データセットである FVQA よりも大幅に大きいものであり、現時点において世界中で最も大規模かつ高品位の知識ベースの VQG データセットである。この課題は、物事の関係性にかかわる知識獲得の試みであり、獲得した知識は直接的に分からないものを理解する Zero-shot 学習にも利用可能である。これはマイルストーンであげた few-shot 学習よりも大幅に困難な課題であることから、当該年度のマイルストーンを超えた成果と言える。

課題推進者：原田達也（東京大学）

研究開発課題 2：意味理解コーパスの研究開発

（1）意味理解コーパスの開発に向けた大規模データの入手と整備

ここでは問診などの文字情報だけでなく、様々な情報を用いた支援を可能にする複数のモダリティ情報からなる意味理解コーパスを開発のための大規模データの入手と整備をおこなった。分野としては計画当初に予定していた放射線科などの他にうつ病等の心療内科的疾患も対象の一つとして考えている。うつ病等の心療内科的疾患の診断は、患者との対話が重要な要素を占めており、医師が行っている対話を CA が代わりに行うことができれば、医師の負担を大きく減らすことが可能となる。また、心療内科的疾患は、昨今のコロナ禍も影響し、社会的にも大きな問題となっており、この分野を支援する CA の開発は社会的にも重要である。そのため、日本心療内科学会と協力し、データの入手するための倫理審査の手続きを進めている。また、当初予定していた放射線科の症例データの抽出も並行しておこなっており、1,000 症例以上の患者の基礎情報の抽出は完了した。

（2）画像情報による少量のアノテーションデータを利用した診断システムの開発

医用データを扱う上での大きな問題の一つがアノテーションコストである。医用画像のアノテーションは、高精度化のためには曖昧な境界を統一されたルールのもとで実施するための時間的なコストや、専門的な知識をもつ人材を確保するためのコストといったものが、一般画像のアノテーションと比べると非常に高い。この問題は本課題で取り組む診断支援システムの開発においても例外ではない。そこで、画像情報を効果的に利用して、このアノテーションコストを削減するための診断支援システムの開発に取り組んだ。アノテーションといってもさまざまなものがあり、今回は所見文中に存在する疾患に関する情報を抽出して検出するタスクに取り組んだ。所見文中の単語同士の依存関係の木構造を獲得して幅優先探索を行うことで、疾患に関する位置の情報を抽出することに成功し、抽出したものの中で正確に位置を抽出したものは 98%と非常に正確に抽出することができた。

課題推進者：黒瀬優介（東京大学）

研究開発課題 3 : 継続学習と記憶の研究開発

(1) 短期および長期記憶のメカニズムにおける基盤研究開発

基本的な相互作用の確認について、以下の分野を探索することで、基礎の構築を行っている。具体的には、人工的知覚の恒常性、長期記憶のナレッジグラフ、短期記憶と長期記憶の相互作用を利用した記憶の記録と検索のメカニズムの設計、脳の前頭連合野の前頭眼窩 (FEF) 領域のシミュレート、短期の空間記憶の設計である。また、長期記憶から短期記憶を探索するための記憶検索手法を設計した。特に、ニューラルネットワーク内の知識を直接解読する研究では秒単位で記憶を抽出することが可能である。検証テストの際、長期記憶に転送する時間は 3 分程度である。不要な情報は、新しい双方向埋め込みベースの生成モデルで 2 時間以内に忘却された。本研究の結果は International Conference on Computer Vision (ICCV 2021), The British Machine Vision Conference (BMVC 2021), IEEE Journal of Biomedical and Health Informatics (JBHI)などの国際会議、論文誌で採録されている。

(2) 自然言語処理技術や医学関連の研究領域を直接サポートする研究開発

自然言語処理と医用画像タスクの記憶システムによるサポートについて、短期記憶メカニズムを利用し医療タスクをサポートする研究開発を行った。実験結果より、3D 腫瘍のセグメンテーションと 2D 胸部 X 線の分類の両タスクにおいて、20%以上の性能向上を達成している。さらに、疾患や病変の関係性に関する医療特化の記憶を設計し、胸部 X 線タスクをサポートする研究開発を行った。

(3) 人間の個性をシミュレートする記憶メカニズムの研究開発

今年度は、ゲーム依存症の人格記憶をシミュレートする研究開発を行った。この記憶システムは CA ユーザーを対象としているため、ゲーム依存症に悩む人を代理ユーザーとして想定している。現在、インターネットゲーム障害 (IGD) の人格記憶を構築し、健常者の感情反応に 50%程度類似していることが確認されている。このモデルに基づいて、後期研究で健常者のモデルを構築することができる。

当初の計画では予見し得なかった進展を生んでいる。PM 石黒浩を訪問した際、記憶のためのニューロン表現についての議論から、この直交表現をシミュレートする人工知能アーキテクチャを設計し、ニューロンネットワークを軽量かつ競争力のある性能にすることのインスピレーションを得ている。また、このモデルをもとに、圧倒的な低消費電力を実現する人工知能チップの設計を考慮した。これらの 2 つの研究は、今年、Association for the Advancement of Artificial Intelligence(AAAI)というトップレベルの AI 国際会議で採録されている。

課題推進者 : Gu Lin (理化学研究所)

研究開発課題 4 : 因果推論と予測機能の研究開発

(1) 複雑な非線形性を有する時系列間の関係性のモデリングに関する研究開発

時系列予測モデルを学習するには予測された系列が真の系列と似通うように、系列同士の離れ具合を損失関数として定式化し、その損失を小さくするようにモデルを学習していく必要がある。従って予測モデルの性能は損失関数の性質に依存する。我々は行動の時系列予測における、各行動に費やされる時間は可変長であり多様である、また行動の起こる順序に意味がある、という性質に着目し、可変長系列が予測できるように生成モデルと、固定長時系列間の距離指標である dynamic time warping の拡張を行った。具体的には生成モデルは予測される将来の行動とともに時系列の終了確率を生成し、dynamic time warping による時系列間のマッチングの損失とともに系列の終了確率が真の系列と一致するように損失を

足し合わせる。

我々は行動カテゴリからなる可変長な時系列データセットを作成し、実際に予測モデルの性能評価を行った。結果的に通常の最小 2 乗誤差を使うモデルに対しては系列長予測誤差は 3 割程度減少、dynamic time warping 損失は 7 割程度減少した。Dynamic time warping を用いた別の手法に対しては性能は同程度なものの計算速度の向上がみられた。

また、高効率なデバイスである Spiking Neural Network において時系列予測を行うための時系列生成手法を提案した（研究開発課題 1（原田）との共同成果であり、時系列モデリング部での貢献）。Spiking Neural Network においては、各層の入出力が 2 値の時系列になるという特徴を持っている。我々は確率的生成モデルである variational autoencoder の生成に用いる潜在変数分布を別の Spiking Neural Network を用いることで 2 値の時系列の確率分布としてモデル化し、さらに時系列確率分布間の正則化項を KL 距離の代わりに maximum mean discrepancy を用いることで全モジュールを Spiking Neural Network により学習可能なモデルを提案した。

MNIST の画像データを時系列に符号化して行った予測実験により、通常の Neural Network を用いた時に比べて 3 割程度の予測誤差の減少と 3%程度の inception score の向上がみられた。また計算量において積の計算回数は 1 桁程度減少した。

課題推進者：椋田悠介（東京大学）

研究開発課題 5：自然言語処理の研究開発

（1）CA 利用に即した汎用意味理解技術の研究開発

CA 利用時に発生する言語による情報のやり取りの状況を精査し、昨年度から CA 利用時の状況に合った意味理解タスクを検討してきた。

まず、自然言語処理の研究分野でこれまでに取り扱われてきた意味理解は、主に新聞記事などを対象として研究が進められてきた。そのため、文法的に正しい文語体で記述され、かつ、事実を不特定多数の他者に誤解なく伝えることを主目的とした文章に適した技術となっている。しかし、CA 利用時に発生する言語による情報のやり取りで扱われる文章は、口語的かつ対話的な文章が多く、基本的に 1 対 1 での会話のように比較的少人数に向けて生成される文章と考えられる。こういった対象となる文章の違いは、利用するデータや方法論の違いとして検討する必要があり、単純にこれまで開発してきた意味理解研究の成果を適用するだけでは良質な結果は得られない。

これまでの自然言語処理研究において似た状況として、新聞記事ではなくブログやショートメッセージサービスのような、いわゆるユーザ生成コンテンツの場合には、その問題特有の現象を考慮しなければ良質な結果が得られないという知見がすでに得られている。このため、ユーザ生成コンテンツを対象とした意味理解の研究も、ウェブが一般化した 2000 年代以降活発に行われるようになった。今回の CA 利用時に特化した意味理解技術の開発は、ユーザ生成コンテンツに対する意味理解研究に近い設定ではあるが、いくつかの重要な相違点が存在する。例えば、CA と操作者、あるいは、CA とユーザは、事前にまたは対話を通して知識の共有がなされていることが仮定されるため、この状況を考慮した意味理解を実現する必要がある。以上の仮定される状況を踏まえ、CA 利用に即した汎用意味理解タスクとして、文章中の単語またはフレーズに対して、CA が知っている概念か知らない概念かを特定する「用語同定タスク」を設計した。「用語同定タスク」は、従来から取り組まれている固有表現抽出/関係抽出、あるいは、ゼロ照応解析で培われた技術を中心として設計できる。一方で、言語的な外部知識（辞書やシソーラス等）を補助情報として活用することが必須となる。以上のことから、CA 利用に即した汎用意味理解の必要要件として、(i) CA 自身が知っている、或いは、知らない用語や概念を把握できる仕組みがあること、(ii) CA と操作者、あるいは、CA とユーザが事前にまたは対話を通して行なった知識の共有に基づく省略や代名詞化の外界照応問題を解決する仕組みがあること、(iii) 外界照応を解決するために、その場で与えられる文章のみならず、外部知識を持っている前提で意味理解を実行できる仕組みを持つこと、の三点を取り上げ、本課題において解決すべき要件とした。

ここで設計された用語同定タスクに対して、実験用データを作成した。マイルストーン設計当初は評価データのみ 3,000 文を作成する予定であった。しかし、人手でチェックする作業コストを、事前に自動で候補を列挙することによって大幅に削減することに成功した。そこで、削減した作業コスト分を更なるデータ作成に充てることとし追加で 12,000 文を作成し、合計で 15,000 文のデータを構築した。よって、当該年度のマイルストーンを大幅に超えて達成した。

(2) マルチモーダル意味理解の研究開発

前記(1)と同様に、タスクの設計と必要要件について考察した。(1)の場合では、言語情報のみを対象としているが、CA と操作者、或いは、CA とユーザが、視覚情報を代表とする各種センサー情報(暑い寒い、匂い、触覚など)を共有する設定の場合に、必要となる意味理解技術に違いが発生するかを検討した。視覚情報に限定した場合でも、上記「(i) CA 自身が知っている、或いは、知らない用語や概念を把握できる仕組みがあること」に関しては、物体検出した上でそれらの物体の知識を持っているかを検出する必要性があり、言語情報単体における既知および未知情報の管理と、それ以外の既知/未知情報の管理を同時に実施する必要がある。また、「(ii) CA と操作者、あるいは、CA とユーザが事前にまたは対話を通して行なった知識の共有に基づく省略や代名詞化の外界照応問題を解決する仕組みがあること」に関しても、視線により話者が何を指し示しているかを省略するといったことが考えられる。これらのことから、外界照応の先を視覚情報にまで拡張する必要性があると判断した。そこで、(1)で作成するデータに視覚情報から得られる物体などの情報を統合し、新たなマルチモーダルタスクとする。以上のことから、CA 利用に即したマルチモーダル意味理解に関しては、(1)で扱うタスクに対して、まずは視覚情報を追加したタスクを本課題において解決すべき要件とした。

そこで(1)で作成した実験用データに視覚情報(画像情報)を付与した。(1)で作成したデータは合計で 15,000 文となるため、本項目のデータも同様に 15,000 文のデータとなる。つまり、当該年度のマイルストーンは計画当初の目標を大幅に超えて達成した。

課題推進者：鈴木潤(東北大学)