

1 研究目的、研究方法など

本研究計画調書は「小区分」の審査区分で審査される。記述に当たっては、「科学研究費助成事業における審査及び評価に関する規程」（公募要領参照）を参考にすること。

本研究の目的と方法などについて、4 頁以内で記述すること。

冒頭にその概要を簡潔にまとめて記述し、本文には、(1)本研究の学術的背景や本研究の着想に至った経緯、研究課題の核心をなす学術的「問い」、(2)本研究の目的および学術的独自性と創造性、(3)関連分野の研究動向と本研究の位置づけ、(4)本研究で何をどのように、どこまで明らかにしようとするのか、(5)本研究の目的を達成するための準備状況、(6)本研究がどのような国際性（将来的に世界の研究をけん引する、協同を通じて世界の研究の発展に貢献する、我が国独自の研究としての高い価値を創出する等）を有するかについて具体的かつ明確に記述すること。

本研究を研究分担者とともに行う場合は、研究代表者、研究分担者の具体的な役割を記述すること。

（概要）

近年注目を集めている人工知能の分野で Google 社や NVIDIA 社が市場を独占していることが問題となっている。この問題は多くの人工知能プログラムが python 言語で書かれていることが原因となってる。この問題をは python 言語のプログラムを Lisp 言語に変換して実行するプログラムである hyclb の開発をすることで解決できる。提案者が過去に作成した hyclb のプログラムには変換できる python プログラムに制限があったが、この制限をなくし任意のプログラムの変換が可能となることを目指す。

（本文）

図 1 に示すように、提案者の作成した hyclb (<https://github.com/niitsuma/hyclb>) というライブラリは累計で 3 万回以上ダウンロードされた。hyclb は hy 言語というマイナーな言語の機能を拡張するライブラリである。3 万回は、マイナーな言語のライブラリとしては多い回数である。hyclb は h y 言語に互換性がなくなるアップデートが行われたことで公開後 1 年で動かなくなったが、稼働していた 1 年の間に hy 言語本体がダウンロードされた回数は、推定で 10 万回と予測している。本体のダウンロード回数 10 万回と比較して、その拡張がダウンロードされる回数 3 万回は多い回数であると言える。以下では、このように多くのダウンロード数があった背景について説明していく。

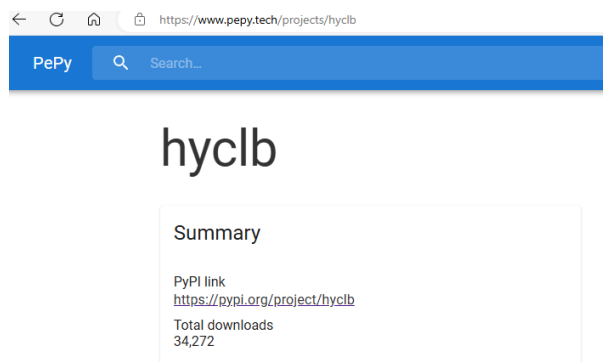


図 1: hyclb 累計ダウンロード数

近年、ChatGPT などの人工知能技術に多くの注目が集まっている。近年発展した人工知能技術はディープラーニングと呼ばれる数値計算に基づく計算で人工知能を実現する技術である。これらの人工知能技術のプログラムの大多数は python 言語を使って開発されている。python 言語は人工知能技術の必須の構成要素である。

python 言語は Google のエンジニアによって開発された言語である。Google の多くの開発で python が使われていると言われている。Google は市場独占をめぐる反トラスト法に反していると疑われており現在、裁判が行われている最中である。

python 言語において Google の市場独占の可能性を疑う事例としては、区間演算と呼ばれる数

【1 研究目的、研究方法など(つづき)】

値計算の基礎的な演算を行うライブラリが長期間存在しなかった事例を上げることができる。約7年前に python 言語で区間演算を行うライブラリが作成された (<https://github.com/taschini/pyinterval>)。しかしこのライブラリは python 言語に互換性のないアップデートが行われ、作成された1年後ぐらいには動かなくなった。それ以降、長期間 python 言語には区間演算という数値計算の基礎的な演算を行うライブラリが存在しなかった。区間演算は基礎的で重要な機能の1つであり、C++やJavaなどの主要な言語の中で、この機能を持たない言語は見つけることができなかった。

このように人工知能技術の必須の構成要素である python 言語に、理由の明確でない変更が行われることは頻繁にある。この変更によって動かなくなるライブラリを Google の市場独占ために選択している疑いは存在する。

なお python 言語に必要なライブラリを動かなくさせる変更があった場合に、変更される前の古いバージョンの python 言語を使い続けられればよいのではないかという疑問を持つ人もいられるかもしれない。しかし python 言語は Linux などの OS の主要機能にも使われおり、古いバージョンの python 言語を使うと、古いバージョンの OS を使わないといけなくなる。古いバージョンの OS の多くにはセキュリティ上の問題があるためネットワークから遮断して使う必要が出てくる問題がある。また古いバージョンの python 言語をでは最新の人工知能技術も動かすことができないため、例えば最新のディープラーニング技術を区間演算を使って検証するといったことも不可能であった。

このような技術の必須の構成要素であるプログラミング言語に、市場独占ための変更が行われることで様々な人が不利益を受けるのを防ぐ方法の1つは、Lisp 系の言語を利用することである。Lisp 系の言語とは Common Lisp や Scheme や hy 言語のような S 式と呼ばれるカッコを主要部品として構成することでマクロと呼ばれる、プログラムの文法そのもののカスタマイズが自由に行える言語のことである。Lisp 系の言語のマクロとは、C 言語の # define のようなプログラムコードの文字列を変更する機能を洗練させたものである。マクロ機能がある場合、例えばもし if then else 文があるプログラム言語から else の機能を省略されて if then しか使えなくなった場合にも、else の機能を簡単に追加することができる。もし C 言語から if then else の else の機能が使えなくなった場合には、非常に多くのライブラリが動かなくなる。else の機能を自分で追加するには C 言語のコンパイラを改造する必要がある。しかし Lisp 系の言語の場合は、5 行程度のマクロのプログラムコードを追加するだけで解決することができる。

hy 言語は、この S 式とマクロの機能を python 言語で使えるようにした python 言語の拡張である。もっと正確に言うなら S 式で記述されたプログラムを、等価な python のプログラムに変換する変換プログラムの集合体である。hy 言語のプログラムは python 言語のプログラムの単なる文字列変換であるため、多くの人工知能のツールやライブラリを、特に変更することなくそのまま利用することができる。

hy 言語が広く普及するならば、python 言語に理由の明確でない変更が頻繁に行われる問題に、Lisp 系の言語のマクロの機能を使って対処できるはずであった。しかし hy 言語には python 言語と同様に、理由の明確でない変更が頻繁に行われる問題があった。例えば `py 2 hy` <https://github.com/woodrush/py2hy> という python のプログラムを hy 言語に変換するライブラリは、理由の明確でない変更で1年以内に動かなくなった。その後も多くのユーザーが同様の変換プログラムが欲しいという要望を上げていた。提案者は、その要望に答えるプログラムを作成した <https://qiita.com/niitsuma/items/42d0b3a60189a592f472>。しかし、これも数ヶ月で動かなくなる変更が hy 言語に行われた。

hy 言語に python 言語と同様に、理由の明確でない変更が頻繁に行われる問題は、元になっている python 言語が変更されることが原因である可能性も考えられたが、それだけは説明しにくい変更も頻繁に行われていた。この問題を解決する手段として提案者が考えた解決策が `hyclb` ライブラリの作成であった。

`hyclb` は hy 言語を Common Lisp に変換してしまう拡張である。この拡張を使うこと

【1 研究目的、研究方法など(つづき)】

で Common Lisp のよく知られたライブラリやプログラムを hy 言語を経由して python 言語上使用ができるようになる。Common Lisp は Lisp 系の言語の代表的な言語であり、マクロが使えるだけでなく、理由の明確でない変更はほとんど行われない安定した言語である。hy 言語にはマクロ機能があるため、このマクロ機能を使って Common Lisp のコードが hy のコードに変換される、というのが hyclb の主なアイデアである。しかし Common Lisp のすべての機能を実装するには時間が足りなかったため、特に有名な Common Lisp のコードである On Lisp(<https://www.asahi-net.or.jp/~kc7k-nd/onlispjhtml/>) の特に有名な関数やマクロが使える範囲に留まっていた。本プロジェクトでは Common Lisp の大部分の機能を hy 言語を経由して python 言語で使えるようにするのが目標の1つである。

hy 言語に理由の明確でない変更が頻繁に行われても、安定した Common Lisp への変換プログラムがすぐに作られるという理由から hyclb はマイナーな言語の拡張としては多くの回数ダウンロードされた。しかし hy 言語の変更を追いつけることが難しくなる、研究が主体ではない職場への転職などの理由により開発は中断してしまい現在に至る。

本研究では hyclb をさらに発展させたライブラリを開発する。hyclb を使うことで、例えば pytorch などの人工知能の分野で非常に利用されている Python のライブラリと On Lisp のマクロを有機的に組み合わせるといったことが可能であった。本研究では、hyclb では実装できなかった残りの Common Lisp の機能を追加で実装することで、さらに様々な Common Lisp の機能を Python のライブラリと組み合わせて利用できるようにする。作成されるライブラリは様々な応用で利用可能であるが、特に自然言語処理への応用で便利となるようなライブラリにする予定である。

研究代表者は自然言語処理の研究も行ってきた。近年 ChatGPT などの自然言語処理に人工知能を応用したアプリケーションが話題である。これら自然言語処理などの人工知能のプログラムはほとんどが Python 言語で書かれている。Lisp 系言語を python で書かれた自然言語処理プログラムと組み合わせて使うことで、どのくらい便利になるのかの検証をしながら、開発を行っていく。

Google は反トラスト法の問題で裁判が行われている最中である。Nvidia も同様に反トラスト法の問題で調査が行われている最中である。Nvidia に関する問題も、前述の人工知能のプログラムはほとんどが Python 言語で書かれていることが原因となっている。これらの人工知能のプログラムは Python 言語の人工知能のライブラリを経由して Nvidia のチップを使うことが必須となっている。そのため、この問題も前述の Lisp 系言語を活用することで解決できる。人工知能の Python プログラムを hy 言語などの Lisp 系言語に自動変換することができれば、Nvidia のチップ以外の任意の環境で実行できるようにすることが可能となる。

【 1 研究目的、研究方法など（つづき） 】

2 応募者の研究遂行能力及び研究環境

応募者（研究代表者、研究分担者）の研究計画の実行可能性を示すため、(1)これまでの研究活動（主要な研究業績を含む）、(2)研究環境（研究遂行に必要な研究施設・設備・研究資料等を含む）について2頁以内で記述すること。

「(1)これまでの研究活動」の記述には、研究計画に関連した国際的な取組（国際共同研究の実施歴や海外機関での研究歴等）がある場合には必要に応じてその内容を含めること。また、研究活動を中断していた期間がある場合にはその説明などを含めてもよい。

(1) これまでの研究活動

本研究は、Lisp 系システムと python 言語で書かれた人工知能ライブラリを融合することで、特定の企業の市場独占を解決できることが目的であった。特に近年重要となっている ChatGPT などの自然言語処理への人工知能応用での特定の企業の市場独占を解決するのが望ましい。そのため、開発は自然言語処理への応用を中心に行っていく予定です。

研究代表者は自然言語処理に関係する研究として以下の研究を行ってきた。

- 文章の印象が良いか悪いかを分析する問題は、感情極性推定と呼ばれている。感情極性推定を Neural Network で行う手法を開発した [4,5,3]。この開発した手法は感情極性推定の正確さと計算速度で従来手法を上回る。
- word2vec は自然言語処理の基礎となる技術であるが、word2vec がカテゴリデータの主成分分析として知られる Correspondence Analysis(数量化 III 類) と等価なことを示した [6]
- ジニ係数は経済格差の指標として良く知られている。ジニ係数が Correspondence Analysis の特殊例であることを示した [7]。これにより word2vec はジニ係数と同じ考え方で作られたことがわかる。

このように単語分散表現 (word2vec) の研究を行ってきており、具体的な自然言語処理の分析対象としては twitter の文章の分析をして観光ルート推薦をする研究も行っている。観光ルート推薦の問題は地図情報というカテゴリデータを分析する問題でもあり、この方向に発展させた研究も行う予定である。観光ルート推薦研究内容が良く分かるデモとして図2の岡山県の観光ルート推薦をするデモプログラムを公開している。図2は岡山駅を 10:00 に出発して倉敷に 18:00 に到着するような観光ルートを生成した例である。デモプログラムは、このようなルートを数分で計算して地図上に表示する。

この観光ルート推薦デモプログラムは以下の研究に基づいて作成されている。

- 「倉敷美観地区」などの岡山の観光スポットの特徴を Yahoo 知恵袋 [3] と Twitter[1] の口コミ情報から分析しユーザーの満足度を計算する。
- 岡山の観光スポットは「倉敷美観地区」以外にも 100 近くあるが、制限時間内 (図2の例の場合 10:00-18:00) に訪問できる観光スポットでユーザーが最も満足できる組み合わせを計算する [1]。図2の例では 1:「鬼の城」2:「サントピア岡山」3:「倉敷美観地区」4:「吉備津神社」5:「蔵の湯」という5つのスポットを順番に訪問するルートが選択されている。
- この観光スポットの組み合わせ最適化問題は (巡回セールスマン問題などと同じ) 計算時間が非常にかかる問題であるが、数分で近似計算をする工夫をしている [2]。

発表文献

1. “Twitter を利用した観光ルート推薦の一手法”: 新井晃平, 新妻弘崇, 太田学, データ工学と情報マネジメントに関するフォーラム (DEIM) (2015)(査読なし)[学生プレゼンテーション賞].
2. “観光ルート推薦のための効率的な制約条件”: 新妻弘崇, 新井晃平, 太田学, 情報処理学会論文誌データベース (TOD), Vol. 9, No. 2, pp. 34–45, (2016). (査読あり).
3. “Yahoo!知恵袋を利用した施設名の曖昧性解消手法の提案”: 中川智也, 新妻弘崇, 太田学, データ工学と情報マネジメントに関するフォーラム (DEIM) (2017)(査読なし).

【2 応募者の研究遂行能力及び研究環境(つづき)】

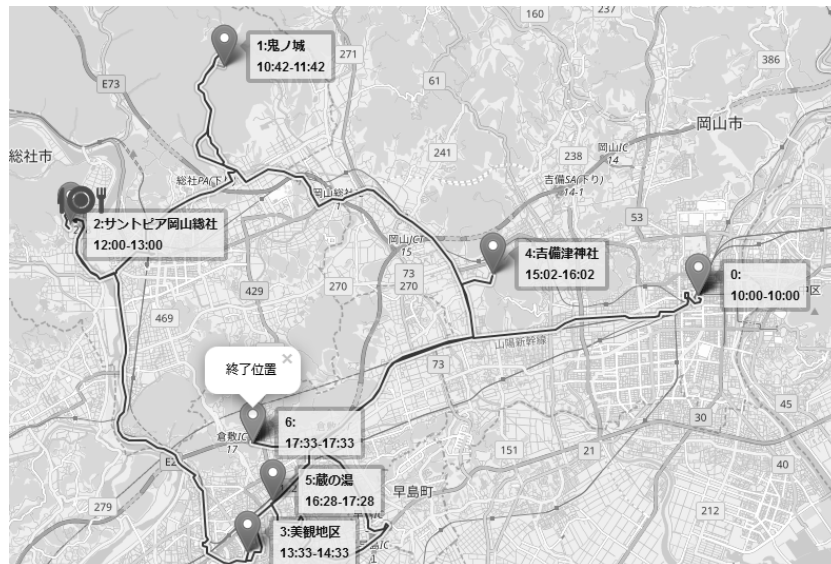


図 2: Twitter を利用した岡山の観光ルート推薦デモプログラム

誰でも利用できる Web アプリとして <http://www.de.cs.okayama-u.ac.jp/~niitsuma/tourrecommendation/tourrecommendations.html> で 2019 年までは公開していた。

KLbibitem “Aspect-Based Sentiment Analysis における Neural Attention の効率的な利用方法”: 朝倉遼, 新妻弘崇, 太田学, データ工学と情報マネジメントに関するフォーラム (DEIM) (2018)(査読なし)[優秀論文賞].

4. “Recurrent Neural Networks on Convolved Word Vectors for Aspect-Based Sentiment Analysis”: R. Asakura, H. Niitsuma, M. Ohta, International Conference on Computational Linguistics and Intelligent Text Processing (CICLing) (2017)(査読あり).

5. “Simple Alignment Sentence Classification for Aspect-Based Sentiment Analysis”: H. Niitsuma, R. Asakura, M. Ohta, International Conference on Computational Linguistics and Intelligent Text Processing (CICLing) (2018)(査読あり).

6. “Word2Vec is a special case of Kernel Correspondence Analysis and Kernels for Natural Language Processing”: H. Niitsuma, M. Lee, arxiv.org/abs/1605.05087 (2016)(査読なし).

7. “Covariance and PCA for Categorical Variables”: H. Niitsuma, T. Okada, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) (2005)(査読あり).

(2) 研究環境

本研究は研究代表者がプログラム開発と実験を行なう。研究代表者が研究指導をしている学生にも本研究のサブタスクを卒業研究の一部として担当させる予定である。

本研究では Neural Network の計算も頻繁に行う予定である。Neural Network には高速な GPU が必要となる。そのため最新の GPU をさらに購入する予定である。また GPU としては様々な企業から発売されているものを比較する。

本研究では自然言語処理の計算を頻繁に行う。自然言語処理には大容量メモリを備えた計算機が必要となる。現在利用しているのは比較的大容量であるがやや容量が足りない場合もある計算機である。これを活用しつつも大容量メモリの最新の計算機をさらに購入する予定である。

3 人権の保護及び法令等の遵守への対応（公募要領参照）

本研究を遂行するに当たって、相手方の同意・協力を必要とする研究、個人情報の取扱いの配慮を必要とする研究、生命倫理・安全対策に対する取組を必要とする研究など指針・法令等（国際共同研究を行う国・地域の指針・法令等を含む）に基づく手続が必要な研究が含まれている場合、講じる対策と措置を、1 頁以内で記述すること。

個人情報を伴うアンケート調査・インタビュー調査・行動調査（個人履歴・映像を含む）、提供を受けた試料の使用、ヒト遺伝子解析研究、遺伝子組換え実験、動物実験など、研究機関内外の倫理委員会等における承認手続が必要となる調査・研究・実験などが対象となる。

該当しない場合には、その旨記述すること。

本研究ではソフトウェアの開発を行う。開発で利用されるソフトウェアは GPL 等の自由なライセンスのソフトウェアと Microsoft Windows , Microsoft Office , MacOSX , Matlab , Mathematica のみである。したがって本研究で利用するソフトウェアに人権保護や法令上の問題は全くない。

4 研究計画最終年度前年度応募を行う場合の記述事項（該当者は必ず記述すること（公募要領参照））

本研究の研究代表者が行っている、令和7（2025）年度が最終年度に当たる継続研究課題の当初研究計画、その研究によって得られた新たな知見等の研究成果を記述するとともに、当該研究の進展を踏まえ、本研究を前年度応募する理由（研究の展開状況、経費の必要性等）を1頁以内で記述すること。

該当しない場合は記述欄を削除することなく、空欄のまま提出すること。

研究種目名	課題番号	研究課題名	研究期間