

# HPCI応募課題のレビュアーを推薦するAIの開発物語

## A Story of AI Development to Recommend Reviewers for the Applied HPCI Projects

高度情報科学技術研究機構  
峯尾 真一

本稿では、革新的ハイパフォーマンス・コンピューティング・インフラ（HPCI）の課題審査において、レビュアー候補選定のために用いている人工知能（AI）と疑似量子アニーラを組合わせたレビュアー割当フレームワークの開発経緯について紹介する。

### 1. はじめに

本年5月8日に新型コロナの感染症法上の位置付けが5類に移行し、やっとかつての日常が戻りつつある昨今ですが、ここでご紹介するAIツールの開発は丁度このコロナ禍中に行われていました。2020年3月13日に新型コロナウイルス対策の特別措置法が成立、同年4月7日に東京等7都道府県に緊急事態が宣言され、4月16日には対象が全国に拡大しました。そこで、2020年を紀元としてBC（Before Corona）とAD（After Declaration of Emergency）の年号を用いてHPCI [1] に応募された課題のレビュアーを推薦するAIツール開発の歴史を振り返ってみようと思います。

その中で申請課題の評価を行うレビュアーはどのように決まるのかという課題選定の舞台裏の一端もご紹介できれば幸いです。

### 2. 1BC年8月～きっかけは課題審査委員会委員長

申請された課題の評価は、レビュアーによる採点の後に、課題審査委員会がその結果を審議して決定されます。そのレビュアーの選定は、年度毎に委嘱される日本全国の研究者のレビュアー候補から、RISTでHPCIの利用支援を担当する職員が申請課題の内容を吟

味して行っていました。

この年8月28日の課題審査委員会において、RISTが選定したレビュアーの妥当性をどの様に担保するのかという問題提起があり、その後の検討の中で、同年11月20日、課題審査委員長から「どの位の作業負担かわかりませんが、レビュアーのアサインは機械学習などデータサイエンス技術を使って90%程度まで自動化できないものですかね？RISTさんの種々の仕事にはかなり利用効果がある気がしていますが…」とのご意見を頂きました。

その頃RIST共用促進部で課題選定に携わっていた私は、以前読んだ本 [8] を参考にして機械学習を試してみることにしました。これが全ての始まりです。

### 3. AD1年1月～機械学習モデルは苦肉の策

年が明けて、まずは機械学習モデルをどう作るかを考えました。素直にモデル化するなら、過去の課題情報と実際に割当てられたレビュアーを対にして訓練データとすることになります。しかし、レビュアー候補は年度毎に変更されるため、学習済みモデルを未来に候補となるレビュアーに適用することができません。苦肉の策としてレビュアーの方を分類するモデルを作ることにしました。

しかし、そもそもの問題として、レビュアーの分類モデルで課題を分類することは、例えば100人の顔を覚えさせたAIに101人目の顔を見せて100人の内の誰かを答えさせることになるという矛盾を抱えることとなります。

#### 4. AD1年2月～AI使い隊の発足

レビュアーを分類するための訓練データとして、レビュアー委嘱時に自己申告で入力された専門分野を示すキーワードを使うことができますが、数が足りません。またAIの結果を客観的に評価する仕組みが必要です。そこでRIST内の専門家に支援を頼むべく、利用支援部長に検討チーム（通称「AI使い隊」）の発足を提案し、利用支援部、産業利用推進部、共用促進部からメンバが集まりました。この時の私のメールには、「とりあえずパッケージとしてTensorFlowをダウンロードしてテストしますが、（私見では）レビュアー選定におけるAI活用についてはデータの性質上、難しいと思っています。むしろ単純な支援ツールの方がよいかもしれません。」と

あり、自信の無さが覗えます。（TensorFlowについては [2] 参照）

#### 5. AD1年3月～レビュアーの分類モデル誕生

AI使い隊は、当初6名で発足し、AIの評価のためにAチーム（人間の判断でレビュアー選定）とBチーム（AIを訓練してレビュアー選定）に分かれて互いに影響を与えないように結果を出すことにしました。過去の記録から110人のレビュアーを選び、AIの分類モデルを隠れ層2段(300,100)のMLP (Multi Layer Perceptron) で作成しました。ディープラーニングとしては小さなモデルですが、先行例 [8] ではMNISTデータセット（手書き数字の認識）問題で十分な性能を発揮します。評価用の課題データも過去の申請課題98課題を用いました。

ここで採用した機械学習の仕組みは単純です。手順は、①レビュアー毎の特徴を示すキーワードの組合せを作成し、②全てのキーワードを網羅するキーワードリストに対して、レ

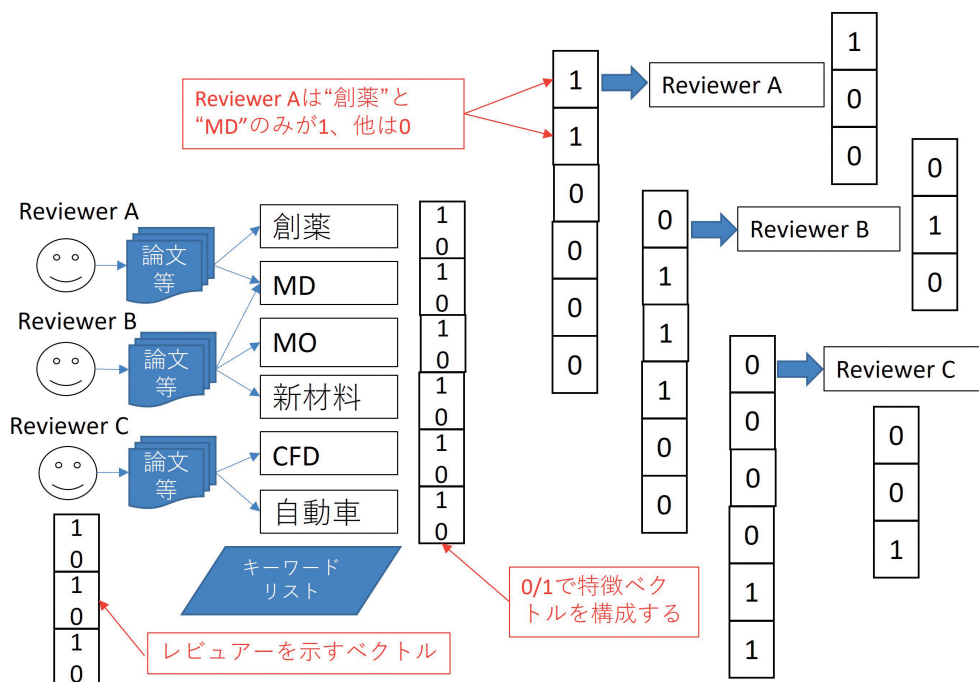


図1 AI分類モデルの訓練データ

ビューアーID毎にキーワードの有無により1(有)と0(無)を用いた特徴ベクトルを作成(図1)、③それを入力とし、レビューアーIDを正解(ターゲットラベル)として訓練します。(図2)

こうして訓練された分類モデルに対し、①申請課題の文章データと各キーワードのパターンマッチングを行い、②課題の特徴ベク

トルを作成して入力すると、③出力の合計値が100%となるsoftmax関数によりレビューアーIDに対応した課題の適合率が求められます。(図3) この結果を課題毎の一覧としてAIのレビューアー推薦表としました。

振り返れば5 BC年12月にResNETが発表[9]されて一挙に100層以上の深さに達するネットワークが出現し、現在のChatGPT

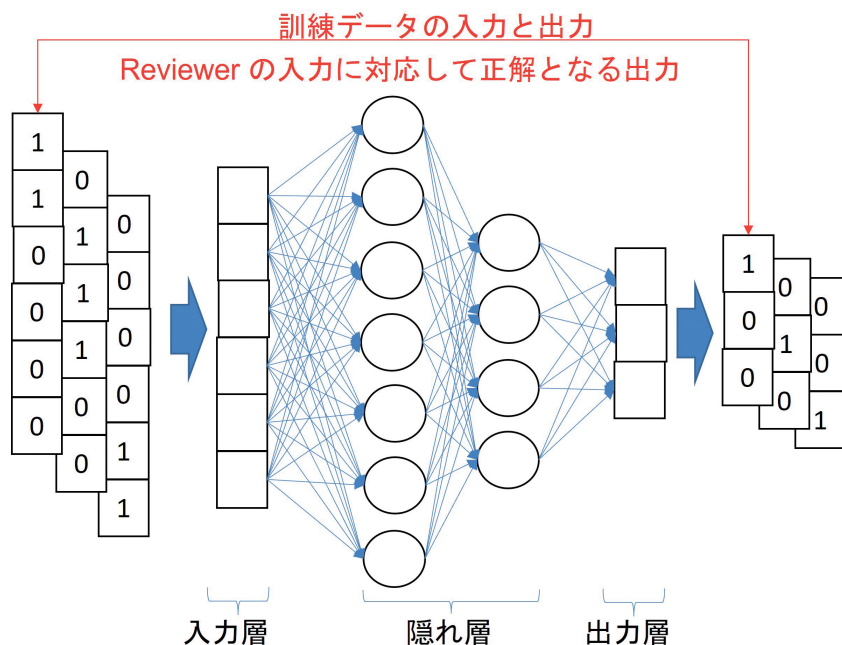


図2 AI分類モデルの訓練

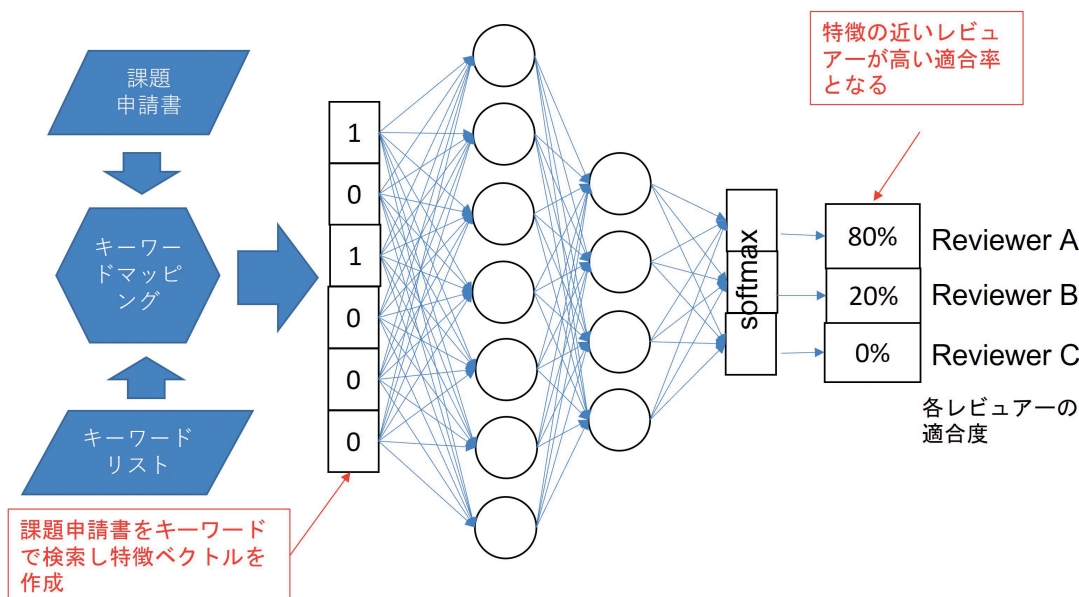


図3 AI分類モデルの利用

に繋がっていく研究開発競争のただ中にいた訳ですが、この時我々は最も単純なモデルでAIの可能性を探っていたこととなります。

## 6. AD1年3月～最初の冴えない結果

訓練データは、各レビュアーに対してBチームがwebや論文から追加したキーワードリスト1組です。(ユニークなキーワード数1026、レビュアー毎には最少8、最大34)この様に訓練データが少ない場合、機械学習ではすぐに過学習と呼ばれる訓練データに特化したモデルができてしまい、他のデータに適応できない(汎化しない)モデルとなります。(図4)

そこで、当初は過学習防止のため、訓練を途中で打ち切ったモデルを作成しましたが、課題の分類が全くできません。敢えて過学習をさせたモデルを使うことにしました。またAI利用(AI法)と並行して、課題申請文中のキーワードの一致数で単純にレビュアーを選ぶキーワード法も試みました。結果の評価は、AI法やキーワード法の推薦するレビュアーの上位15人に、人間(Aチーム)が決めたその課題の専門分野のレビュアー(最少

15人で分野毎に異なる)が何人入っているか(A:B一致率)を測定して行いました。

A:B一致率の結果は、キーワード法の60%に対し、AI法は47%(5課題が0%)。AI法の特徴(欠点)は複数の課題でキーワード一致数が0の適切でない同一レビュアー(複数分野に渡る8語だけのキーワードを持つ)を上位に推薦してくることです。また訓練する度に挙動の異なるAIの分類モデルが生成されるため、どれを選ぶかが問題となります。マーフィーの法則「どうしても直せない問題点は、それを特徴とせよ」に従い、一致数0のレビュアーを作為的に入れて、そのレビュアーが上位に推薦されるモデルを棄却するという使い方もあるのではないかと冗談を言っておりました。(図5)

しかし、結局キーワード法もまだ実用レベルではないため、RIST内の各研究分野の専門家を追加してレビュアーのキーワードの見直しを行い、また、秋の定期募集での試験的導入を目指すことにしました。

この年の4月に新型コロナウイルスの緊急事態宣言のために事務所は閉鎖され、完全な在宅勤務となりました。しかし、在宅勤務に

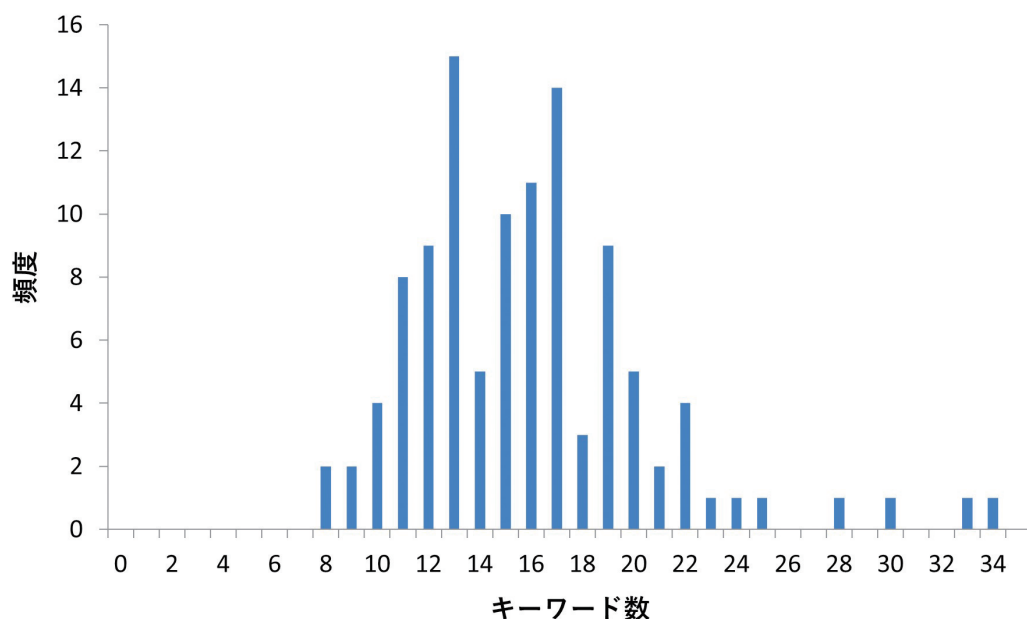


図4 レビュアー毎のキーワード数

より利用可能となった通勤時間は、AIツールの開発にとっては有効であったと思っています。

### 7. AD1年4月～キーワードに問題あり

これまでの結果から、バイオ・ライフ系と物質・材料系の課題が区別できない等の問題が発覚し、さらにRIST内の専門家を増員してキーワードを見直しました。レビュアー当りのキーワード数の目安を20～30とし、各分野の専門家にAIの分類モデルを配布して効果を見ながらキーワードの追加をしてもらうことにしました。

またキーワード法についても色々なアルゴリズムが提案され、並行して検討することになりました。

AI法にとってもキーワードが要となるの

で、まずは適切なキーワードを集めることにし、論文等の情報から自然言語処理を用いてキーワードを自動抽出する方法を検討することになりました。

但しキーワードを用いる方法には潜在的な問題(表1)があることが分かって来ました。

### 8. AD1年8月～AI法のブレークスルー

AI法に新しいアルゴリズムを導入しました。レビュアーの分類用キーワードセットから課題の分類モデル用の訓練データを作るという案です。すなわち、そもそも個々の課題申請書でパターンマッチングするキーワードは少ないので、レビュアーのキーワードセットからランダムにn個(サンプリング数nは当初8)を取り出して機械学習の入力とし、課題申請データをシミュレーションします。

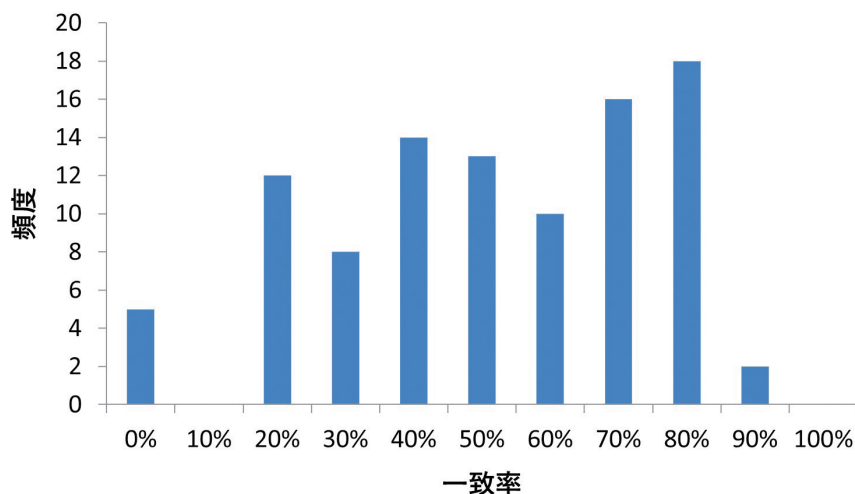


図5 AI法のA:B一一致率

表1 キーワードの問題点

項目	例	問題点
カタカナの表記ゆれ	「ナヴィエ・ストークス」と「ナビエストークス」	文字の違い、「・」の有無
複合語の扱い	「磁気乱流」	「磁気」+「乱流」+「磁気乱流」の3つを登録するべきか否か
同義語の扱い	「分子動力学」と「MD」	正式名称と略語、英語と日本語等の同義語
誤植の扱い	「観測ビクデータ」と「観測ビッグデータ」	「ク」と「グ」の違い

表2 AI法とキーワード法の比較 (AI:AI法、KW:キーワード法)

研究分野	課題数	AI一致率	KW一致率
気象・海洋	6	94.4%	94.4%
流体・熱	17	54.9%	52.5%
構造解析・材料力学	2	72.2%	66.7%
核融合	0		
原子力	0		
宇宙・天文	10	70.0%	67.5%
プラズマ・超高層物理	5	73.3%	77.8%
生物科学・創薬・医学	15	87.4%	88.9%
物性物理	5	45.0%	41.7%
材料化学	19	40.0%	47.4%
地震・地球・惑星形成	0		
素粒子・原子核物理	16	95.1%	94.2%
数値/統計計算・データサイエンス	3	40.0%	26.7%

これで自然な課題の分類モデルを作ることができました。

この方法は、過学習を防ぐための正則化手法のひとつである「半教師あり学習」と類似しており、過学習抑制効果があると考えられます。

新しく追加されたキーワードを組み込んだ結果は、A:B一致率でAI法が47%から66%に、キーワード法が60%から68%に向上しました。AI法は大きく改善しましたが、未だキーワード法に対する優位性はありませんでした。但し、キーワード追加の効果はAI法の方が格段に大きく、さらなる進化によりキーワード法を凌ぐ可能性は大きいと感じました。

この後、正解率が低い構造解析・材料力学系レビュアーのキーワードの改善について専門家に支援を求めて見直した結果、キーワード法の改善は見られませんでした。AI法のA:B一致率は68%とキーワード法と互角になりました。但し研究分野毎にそれぞれ特徴があり、正解箇所が完全に同じにはなりません。

(表2) キーワード法はパターンマッチングの一致数が同じレビュアー間で順位が決められないという欠点があるため、AI法を主とすることにしました。

また、AI法とキーワード法を組合せて使うと、A:B一致率を81%まで高められることが分かりました。(表3) 各課題でキーワード一致数1位のレビュアーの専門分野を調べ、AIのレビュアー推薦表で、課題毎にその専門分野のレビュアー全ての適合率に+1のバイアスを与えて推薦表を再出力するというものです。しかし、この方法では分野融合的な課題にとって不適切なバイアスとなることから採用しませんでした。

この成果については9月の課題審査委員会に報告し、次回定期募集課題において、AI

表3 A:B一致率の比較

方式	人間との一致率平均
キーワード法(KW)	68%
AI法(AI)	68%
AIとKWの組合わせ	81%

法から得られるレビュアー推薦表を補助的に用いてレビュアーの割当を実施することになりました。また、委員からの技術的な質問に対応するべく論文としてまとめ、この年のAXIES（大学ICT推進協議会）年次大会で発表しました。[10]

## 9. AD1年11月～AIの初陣

HPCIの定期課題募集は、A期（前年秋に募集し4月利用開始）とB期（春に募集し10月利用開始）の年2回あります。AIの初陣はこの年の11月、翌年度A期課題の審査となりました。

AIは全申請課題の内、英語で記述された申請書（以下、「英語課題」という。レビュアーのキーワードが日本語のため英語課題には対応できません）と産業利用課題（レビュアーは産業利用専門として別途選ばれる）を除いた133課題に対して、114人のレビュアー候補から推薦順位を計算しました。これを参考にして、人間が組合せを考えます。制約条件は、課題とレビュアーの専門分野が一致していること、課題当りのレビュアー数3人（資源提供機関からのレビュアーを除く）、相対評価を行うためにレビュアー当りの課題数の目安を5以上10以下としました。

結果は、各レビュアーに割当てた課題の採点を依頼した際の、利益相反申告による再割当が多発してしまいました。また、課題審査委員から、レビュアーの専門分野を複数定義すべき、課題毎のレビュアー数を増やすべき、とのご意見を頂きました。

## 10. AD2年1月～AIツール改善の模索

正月明け早々に、これまで試して来たAIを含むツールの改善に取り組みました。まずは、利益相反の多発に対して、過去の課題審査時の利益相反記録を取り込み、AIのレビュアー推薦表で、課題参加者とレビュアー候補を突き合わせ、過去に利益相反となった組合

せのレビュアー適合率を-1に置換えました。（この様な調整済み適合率を適合度と呼びます）適合率は0.0から1.0の間なので、AIの推薦順位は最低位となります。さらに利益相反条件となる他の条件も含めて、利益相反リストを作成することにしました。

また課題と同じ分野のレビュアーを優先するために、AIのレビュアー推薦表において、課題の分野と同じ専門分野のレビュアーに対して+1のバイアスを与えることにしました。このため、レビュアー候補と申請課題について、予め人間が専門分野に分類しておく必要がありますが、この分類はかねてから人間がレビュアーを割当てるために科研費の中区分に沿って実施しているため負担は増えません。

課題当りのレビュアー数を増加することについては、当時のレビュアー候補数ではレビュアーが不足することが明らかであるため、慎重に検討を進めることになりました。しかし、もし課題当りのレビュアー数を増やした場合、組合せ爆発が起きて人間の手に負えなくなることが目に見えており、何らかの対策が必要となりました。例えば5課題に対し、レビュアー候補が4人の簡単な場合でも、課題当り2人、レビュアー当り2以上3以下の課題分担として組み合わせは1,860通りにもなります。実際には申請課題数・レビュアー数ともに100以上となります。

## 11. AD2年4月～量子アニーリングへの期待

組み合わせ最適解を求めるために、新たな手段が必要となりました。原理的に量子アニーリングが使える可能性があることは分かっていたのですが、量子アニーリングマシンで扱うことができる問題規模は、例えば最新のD-Waveシステムでも5000量子ビットに過ぎません。実際の問題は200課題x200人レビュアー＝4万量子ビットの規模となります。

しかし丁度この頃に（株）Fixstarsから

GPUサーバを量子アニーリング用にチューニングした、疑似量子アニーラのクラウドサービスが始まりAmplifyというライブラリがリリースされました。[3] 無料で10秒の試行ができるので、半信半疑で前章の5課題4レビュアーの組合せ問題を解いてみると、全組合せを計算して解いた答えと一致しました。さらに前年度課題のデータで試すと300msで尤もらしい答えが返って来ます。そこでこれを導入することになりました。

AIの計算するレビュアーの適合率は、その課題に対するレビュアー間の相対的な値であるため、絶対値として課題間で比較することはできません。そこで、適合率（実際には調整済みの適合度）を課題毎のレビュアーの優先順位に変換し、全割当レビュアーの優先順位の合計を最小にする目的変数を定義しました。入力データは1からレビュアー数までの整数が並ぶ行列となり生データではないので、クラウドサービスを使う上でのセキュリティ上の懸念も無くなります。

Amplifyのアルゴリズムは次の様なQUBO (Quadratic Unconstrained Binary Optimization) モデルを使うものです。(参考資料 [7])

M件のプロジェクト(申請課題)があり N人

のレビュアー候補がいる場合の決定変数を、

$$q_{ij} = \begin{cases} 1 : \text{プロジェクト } i \text{ にレビュアー } j \text{ を} \\ \quad \text{割当てた場合} \\ 0 : \text{それ以外の場合} \end{cases}$$

として、目的変数  $y$  を、割当てたレビュアーの順位の合計である

$$y = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} w_{ij} q_{ij}$$

但し  $w_{ij}$  をプロジェクト  $i$  に割当ててるレビュアー  $j$  の優先順位とする。制約条件は、全てのプロジェクト  $i$  について、プロジェクト毎にレビュアーを  $c$  人とする事：

$$\sum_{j=0}^{N-1} q_{ij} = c \quad \text{for all } i$$

および、全てのレビュアー  $j$  について、割当てる課題数を  $0$ 、または  $a$  以上  $b$  以下とすること：

$$\sum_{i=0}^{M-1} q_{ij} = 0 \text{ or } \geq a \text{ and } \leq b \quad \text{for all } j$$

となります。

Amplifyは大変使い易いライブラリですが、レビュアー割当に関して一つだけ問題が

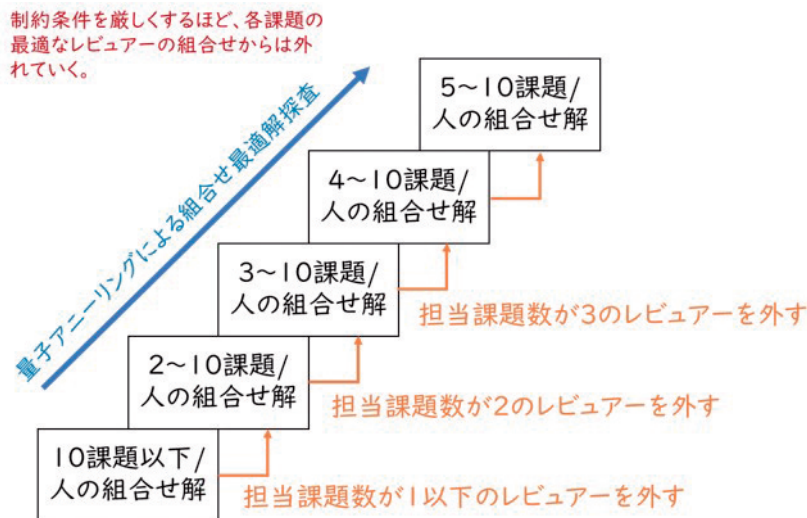


図6 漸近的最適解探索



ありました。レビュアー候補の内、課題に割当ててるのはその一部であるため、レビュアー当たり0または $n$ （最小値）以上 $m$ （最大値）以下の課題という複雑な制約条件となることです。ライブラリではmaxまたは区間条件が指定できますが、区間条件では全てのレビュアーにその条件が適用されてしまいます。

そこで次善の策として段階的にレビュアー候補を絞っていくアルゴリズムを開発しました。(図6)

さらに、この様にして作成した5組の解を組み合わせて、できる限り課題と同じ専門分野のレビュアーを割当てると同じような解を見つけるアルゴリズムを導入しました。この成果は2022年度AXIES年次大会で発表しました。[11]

## 12. AD2年5月～B期課題への試験的適用

これまでの成果を全て取り込み、B期課題申請(51課題、レビュアー候補263人)へ試験的に適用しました。AIは、課題と同じ専門分野で、既知の利益相反が無く、必要な制約条件(課題当りのレビュアー数、レビュアー当りの課題数)を満たし、申請課題内容に近いレビュアーの組合せを提案することができるようになりました。B期は申請数が少ないので人間が割当てすることも簡単にでき、その結果と突き合わせてシステムの機能確認を行いました。

機能的には(誤りも想定内で)大きな問題はありませんでした。AI使い隊において次回A期課題に向けて以下の懸案事項への対応を図ることにしました。

### (1) レビュアー数増加

課題当りのレビュアーを5人とするべく、過去の課題データからA期で想定される申請規模の模擬データを作成し、追加予定のレビュアーで足りるかどうかをシミュレーションして確かめる。また新規レビュアー候補のキーワードは、科研費DB等か

ら半自動的に作成するツールの作成を試みる。

### (2) レビュアーの専門分野の複数化

組合せがさらに多様となるため最適な組合せを人間が評価できないとの意見もありましたが、分野横断的な課題への対応として必要であることが合意されました。

### (3) 産業利用課題対応

これまで産業利用課題は申請数が少ないことからAIツールの対象外としていましたが、課題申請の締切後に、産業利用課題のレビュアーを先に決めてから他の課題種類のレビュアー候補を確定していたため、産業利用課題レビュアーの決定が律速になり作業期間の短縮ができないことが判明しました。そこで産業利用課題にもAIツールの適用を検討することになりました。

## 13. AD2年11月～A期課題でAIツールを本格投入

AIの適用対象は、英語課題を除く135課題です。11月の本番に向けての半自動処理によるキーワード追加の結果、280人のレビュアー候補に対して、キーワードのべ設定数17,403、1人当たり最少24語、最多174語が定義されました。定義数の大きな格差がどの様に影響するのかを調べる時間はありませんでしたが、課題と同じ分野が優先されるバイアスという安全弁があるので、とりあえずこのまま使うことにしました。

MLPのハイパーパラメータである隠れ層のノード数は、入力となる特徴ベクトルの長さ(=ユニークなキーワード数4,945語)に合わせて2,700と900に拡大し、課題申請のシミュレーションのためのサンプリング数を8から16に変更しました。(キーワード数の増加で課題申請書とのパターンマッチング数も増えると期待されるため、またサンプリング数が8のままであると訓練時間が8時間以上必要となることが分かり訓練を加速する必要

があるため)

作業過程でいくつかの問題が顕在化しました。課題の分野推定をする人間の判断基準の個人差、残った英語課題を後から手作業で対応するための煩雑性などです。

課題の分野推定に関しては、次回から専門家を動員してチェックすること、また自動推定の可能性を検討することが決まりました。英語課題に関しては、①日本語のキーワードを英訳したキーワードセットを準備し、申請課題の文書を英語キーワードとパターンマッチングして特徴ベクトルを作成する案と、②申請書を自動翻訳して日本語にするという案が提案され検討することになりました。

#### 14. AD3年5月～B期課題でGPUがほしくなる

新年度のレビュアー候補279名が決まり、キーワードの半自動生成ツールにより、1人当たりのキーワード数は、最少12語、最多278語となりました。(ユニークなキーワード数6,560語) キーワードが極端に少ない2人については機械学習の対象から外し、人間の判断で割当てることにしてAIモデルの訓練を行いました。MLPのハイパーパラメータである隠れ層のノード数は、2,700と900で

同じです。

機械学習はノートPCで実行して来ましたが、ユニークなキーワードの増大により4時間近くかかるようになり、そろそろGPUサーバの出番かと感じました。(図7)

B期課題のレビュアー選定作業は無事に終了しましたが、AIの推薦ではレビュアー当たりの課題数を5以上10以下に近づけるためにレビュアーを集約して割当てるので、適合度から見て必ずしも最適な解とはなっていません。最終的には専門家が確認して必要に応じて再割当する必要があります。AI使い隊の振り返り議論では、今後の課題として、レビュアーの負荷平準化、専門家がレビュアーを再割当する方法の簡便化、英語課題への対応が挙げられました。

この後A期課題審査までに、以下のような数々の試作ツールが誕生します。

- (1) キーワードセットの英訳(13章①の方法を採用)による英語課題対応ツール
- (2) 申請課題の研究分野推定ツール
- (3) Flaskを用いたWebインターフェースによる再割当ツール

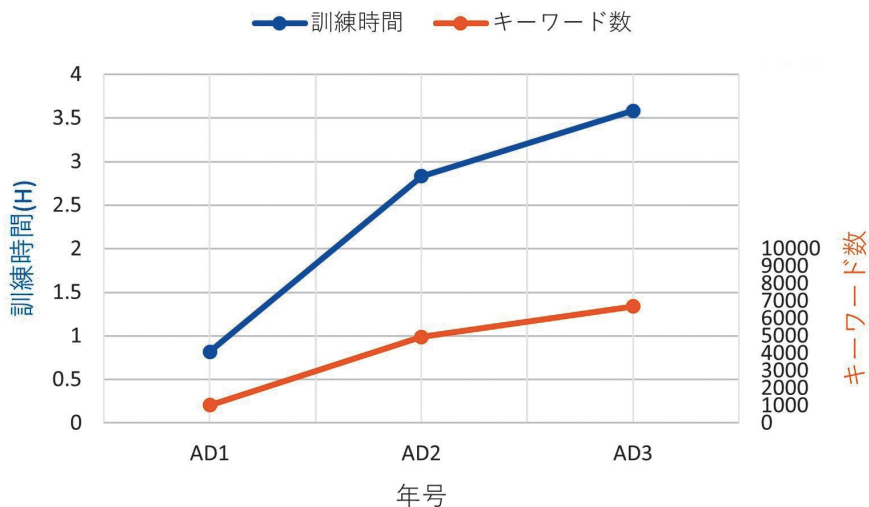


図7 訓練時間

### 15. AD3年11月～A期課題での総決算

今回のレビュアー割当方針についての事前打合せで、レビュアーの負担軽減対策が提案され、過去に担当課題数が多いレビュアーは優先度を下げ、課題の割当を抑制すること、また、利益相反時の再割当の自由度を確保するため、初期の割当をレビュアー当り最大8課題とすることが決まりました。(再割当時、上限の10課題割り当てても良いとする)

課題数178、レビュアー候補者数237、レビュアー当り最少31語、最大268語のキーワード(ユニークなキーワード数6,688語)、産業利用、英語課題、課題の分野推定ツール、Webインターフェースの再割当ツール、レビュアーの負担軽減を全て組み込んだシステムが稼働しました。

その後、レビュアーの負担軽減策で、対象レビュアー候補の適合度を一律1.0(課題と同じ分野で最低の優先度)に置換えたため一部の分野で元々の適合度が消えて再割当がしにくいという指摘があり、今後は1.0+元の適合率とすることにし、全体のバイアスを見直しました。またTensorFlow [2] からPyTorch [4] に書替えてGPUサーバの活用を図ることになりました。

### 16. AD4年～新しい出発

やっとレビュアー割当の一連の工程を全て含むAIツールが実現しました。後はこのプロトタイプを実運用に向けたフレームワークとして完成させることが目標となります。これまで機能追加を繰り返し数珠繋ぎに動作しているコード群を書替えて、データベースを中核として各機能コードを独立させれば、利便性・運用性・保守性を大幅に向上させることができます。(図8)

今後の課題として考えられるのは次の点です。

#### (1) 機械学習モデルのチューニング

現在は最も簡単なMLPですが、モデルの多層化を含むハイパーパラメータのチューニングは全くできていません。DNN(Deep Neural Network)モデルは大きな発展を遂げており、多様な可能性が残っています。

また訓練データについてもキーワードのワンホットエンコーディング(有/無に対応する1/0)から単語の埋め込み表現(単語の相関関係を表現する浮動小数点表現)にすることも考えられます。

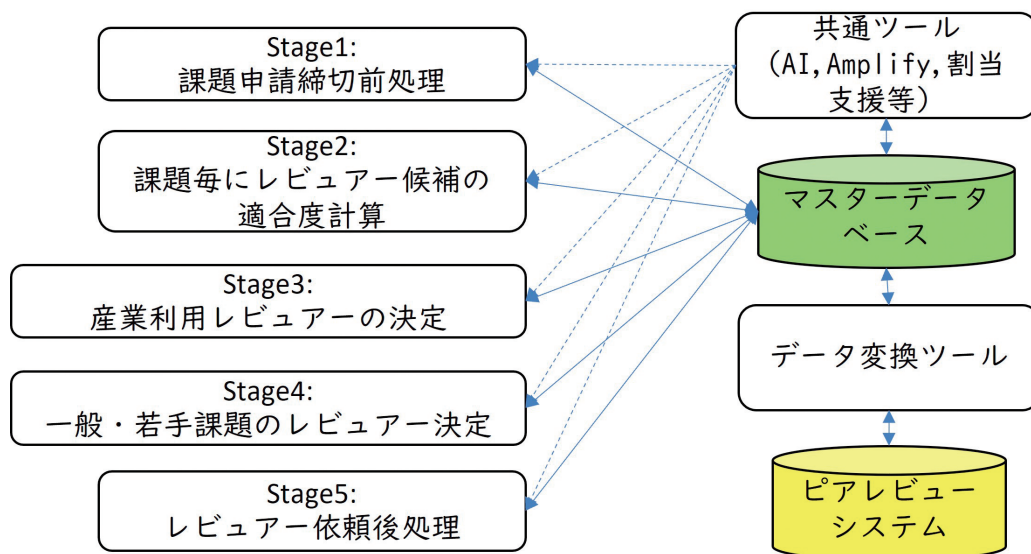


図8 統合フレームワーク

## (2) キーワード生成の問題解決

7章で述べた問題点は、計算科学用語辞書を開発し自然言語処理を用いて解決できると考えます。しかし、根本的に「意味」ではなく「単語」の一致を使う限り限界があるとも言えます。

また半自動生成ツールによりキーワード数が激増していますが、AD4年度のレビュアーに定義されたキーワード6,706語の内、実際に過去6年間の申請課題にヒットするのは2,302語（34%）だけという現実があります。レビュアー当りのキーワード数の格差が10倍近いことと合わせて、今後改善を検討する余地があると考えています。

## (3) DNN以外の新しい方式の検討

RIST内でもキーワード法を改良してヒューリスティックな評価関数によるレビュアー割当を模索する動きがあります。またキーワードに頼らない新たな方法も考えられます。(17章参照)

## 17. 未来の可能性～夢を再び

実はAIツール開発の初期段階で実験してみても断念した方式案がありました。それは自然言語処理による課題とレビュアーの完全自

動分類です。

現在の方法では、キーワードの生成や課題・レビュアーの専門分野決定について人間の判断が入らざるを得ず、工数的・時間的なボトルネックになっています。またキーワードには、7章で述べた問題点が付きまといま

す。キーワードを使わずに処理ができたという思いから、自然言語処理に着目しました。そこで4 BC年にFacebookのAIリサーチラボによって開発されたテキスト分類を学習するためのライブラリfastText [5] を使い、次のような手順を考えました。

- (1) 科研費データベース [6] の審査区分／研究分野」における中区分（65分類）に登録されている論文の「研究開始時の研究の概要」を用いて、fastTextに中区分の分類を学習させる
- (2) レビュアー候補の論文（複数可）を科研費データベースや本人から提供を受けて入手し、fastTextに分類させ、65次元の特徴ベクトルを算出しておく
- (3) 課題申請書に対しても同じ分類を行い、65次元の特徴ベクトルを計算する
- (4) 65次元空間において課題申請書の特徴ベクトルからレビュアー候補の論文を探索

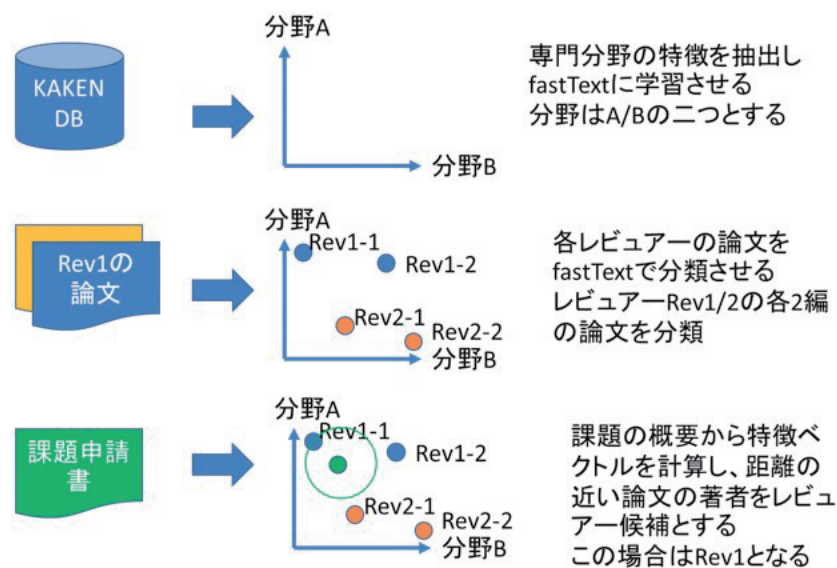


図9 二分野の分類例

し距離の近い順にレビュアーとして割当てる (図9)

この方法なら全てが機械的な操作で完結し専門家の手を煩わせる必要はなくなります。実験は、当時登録されていた

- ・中区分16:天文学およびその関連分野 (49 records)
- ・中区分19:流体工学、熱力学およびその関連分野 (57records)
- ・中区分48:生体の構造と機能およびその関連分野 (69records)

のテキストを日本語形態素解析エンジン MeCab [12] で分ち書きし、中区分のラベルを添付してfastTextを訓練しました。

次に過去の課題申請書から上記3区分に該当する概要テキストを選び、同じように分ち書きした上で分類させてみた結果が表4です。

表4 fastTextの結果

分野	課題数	正解数	正解率
宇宙・天文	10	10	100.0%
流体・熱	17	7	41.2%
生物科学・創薬・医学	15	13	86.7%

分類が簡単な例題を選んだはずですが、課題分類がこの様な状況では見込みがありません。その原因としては

- ・訓練用データ (AD1年8月時点で登録されているものは全て取り込んだのですが) が少なすぎる
  - ・fastTextの日本語処理能力に限界がある
  - ・日本語形態素解析のためのユーザ辞書 (計算科学用語辞書) がない
  - ・課題申請書の概要が短すぎる
- ことが考えられます。

この時は自然言語処理の限界を感じてその先には進みませんでした。ChatGPTの様な強力なAIが出現した現在、再考してもよ

いのではないかと考えています。

## あとがき

本稿でご紹介したAIのレビュアー割当フレームワークのプロトタイプ開発において、AIのモデル評価やレビュアーキーワードの生成については、利用支援部、産業利用推進部、共用促進部の多くのRIST内メンバ (敬称略・順不同: 草間義紀、末安直樹、野口孝明、小久保達信、山木大輔、太田幸宏、山岸孝輝、澤井秀朋、富山栄治、松岳大輔、吉澤香奈子、寺澤麻子、宮本健吾、志澤由久、岡田達夫、塩原紀行、新宮哲、須永泰弘、斉藤精久、河合淳、宮川良一、山本豊、新山美里) の協力が必要でした。その貢献に感謝するとともに、今後若いメンバがAI活用を引き継いで大きく発展させて行くことを期待します。本年 (AD4年) のChatGPTの世界的な大ブレイクを見ても、AIはまだまだ底なしの可能性を秘めていると感じています。

～いつか来る AI世とは聞きしかど 昨日今日とは思はざりしを～

## 参考文献

- [1] HPCIについて (HPCIポータルサイト)、  
[https://www.hpci-office.jp/pages/about\\_hpci](https://www.hpci-office.jp/pages/about_hpci)
- [2] TensorFlow,  
<https://github.com/tensorflow/tensorflow>
- [3] Fixstars Amplify,  
<https://amplify.fixstars.com/ja/>
- [4] PyTorch,  
<https://github.com/pytorch/pytorch>
- [5] fastText,  
<https://github.com/facebookresearch/fastText>
- [6] “科学研究費助成事業データベース”、  
<https://kaken.nii.ac.jp/ja/>

- [7] 田中 宗、棚橋 耕太郎、本橋 智光、高柳 慎一、「量子アニーリングの基礎と応用事例の現状」低温工学 53, p.287-294 (2018)  
<https://doi.org/10.2221/jcsj.53.287>
- [8] Aurélien Géron、下田倫大監訳、長尾高弘訳、“scikit-learnとTensorFlowによる実践機械学習”、(株) オライリー・ジャパン (2018)
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition”, arXiv:1512.03385 (Dec 2015)  
<https://doi.org/10.48550/arXiv.1512.03385>
- [10] 峯尾 真一、小久保 達信、野口 孝明、草間 義紀、“HPCIの課題審査におけるレビュアー選定のためのAI 活用の試み”、大学ICT推進協議会 (AXIES) 2020年度年次大会
- [11] 峯尾 真一、野口 孝明、草間 義紀、松田 佳希、“量子アニーリングを用いたレビュアー割当最適化の検討”、大学ICT推進協議会 (AXIES) 2022年度年次大会
- [12] “MeCab: Yet Another Part-of-Speech and Morphological Analyzer” ,  
<https://taku910.github.io/mecab/>