

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

峯尾 真一, 小久保 達信, 野口 孝明, 草間 義紀

高度情報科学技術研究機構

mineo@rist.or.jp

A Trial of using AI to select reviewers for evaluation of HPCI project application

Shinichi Mineo, Tatsunobu Kokubo, Takayuki Noguchi, Yoshinori Kusama

Research Organization for Information Science and Technology

概要

本研究では、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）の課題審査におけるレビュアー候補の選定にAIを活用する可能性を調査した。その結果、単純なMulti Layer Perceptron(MLP)であっても、人間の判断との一致度を66%まで高めることができた。またキーワード一致法と組み合わせれば一致度は80%を超えたが、分析の結果、現状の人間判断との一致度の定義が解の良し悪しを評価するのに十分でないことなど、今後の検討課題も明らかになった。

本研究の結果から、現状ではAIによってレビュアーの選定を完全に自動化することはできないが、候補者案を作成することは可能であり、特に分野融合的な研究においては有効であると考えられる。

1 はじめに

HPCI[1]の課題審査においては、各課題に対し予め登録されたレビュアー候補の中から適切な人を複数選んで評価を依頼している。しかし様々な分野の多数の応募課題に対し、多岐に渡る専門分野のレビュアーから適切な人を選ぶことは容易ではなく、総合的な知見を備えた人材が多く時間をかけて作業している。そこで人工知能(AI)を活用して応募課題を評価するのにふさわしいレビュアーを探す試みを行った。

2 AI活用の目的

これまでは登録されたレビュアーを、予め科研費の近い分野の小区分をいくつか束ねた区分に分類しておき、応募された課題内容に照らしてその区分のどれに相当するかを判断し、その区分に分類されたレビュアーの中からふさわしいと思われる人を選定して来た。

定期課題募集の時には200件程度の課題毎に、登録されたレビュアー約100人の中から3人以上のレビュアーを選定する。課題毎のレビュア

一数、レビュアー毎の課題数等の条件を満たし最適な組み合わせを求めることは、その選択肢の多さから難度の高い最適化問題となる。さらに課題によっては従来の分野区分に当てはまらないもの、分野融合的なテーマを持つものがあり、画一的なレビュアー選択ができない場合がある。

そこで、AIの活用により、人間の労力を省くとともにより広い範囲での最適解を求めることを目指した。採用したのは、MLPによる機械学習システムにレビュアーの専門分野に関連の深いキーワードの組み合わせを学習させ、各課題に対し最適なレビュアーを推薦させる手法である。

3 機械学習の活用方法

応募課題に対して最適なレビュアーを探すことを、各課題のレビュアー群への分類問題に置き換え、機械学習の手法を適用した。機械学習コードの作成においてはTensorFlowサンプルコード[2]を参考とした。分類器の構成、訓練データの作成、訓練課程、実行過程については以下のとおりであ

る。

3.1 分類器の構成

隠れ層 2 段（1 段目 900 ノード、2 段目 300 ノード）のプレーンな MLP により分類器を構成した。また活性化関数として ReLU(Rectified Linear Units)を、出力層には softmax 関数を用いた。

3.2 訓練データの作成

訓練データの作成手順は以下のとおりである（図 1 参照）。

- ①各レビュアーのキーワードを Web 情報や論文から人力で抽出する
- ②全てのキーワードを集めて重複のないキーワードリストを作成する。
- ③キーワードリストにおいて、各レビュアーの保持するキーワードを 1、保持しないキーワードを 0 で置き換えた 1 次元ベクトルを作成し、入力データとする。
- ④各入力に対する正解が当該レビュアーとなる訓練データの対を作成する。

3.3 訓練過程

上記の訓練データを MLP に適用し、バッチ勾配降下法で訓練を行う（図 2 参照）。

3.4 実行過程

3.2②のキーワードリストの各キーワードについて、応募課題の課題申請書に記載された「課題名」、「課題概要」、「キーワード」を対象に検索し、文字列の一致がある場合は 1、無い場合は 0 として 1 次元ベクトルを作成する。この 1 次元ベクトルを入力として 3.3 で訓練した MLP に分類させる。結果の出力は、各レビュアーに対する適合度を成分に持つ 1 次元ベクトルとなる。（図 3 参照）。

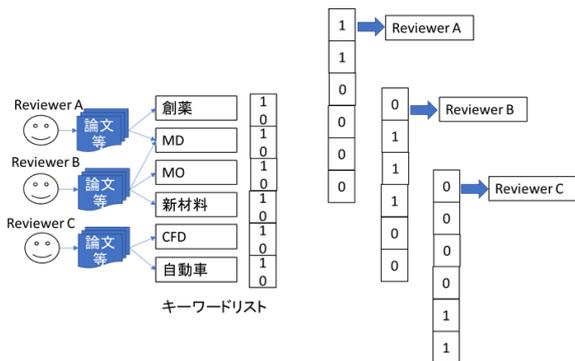


図 1. 訓練データの作成

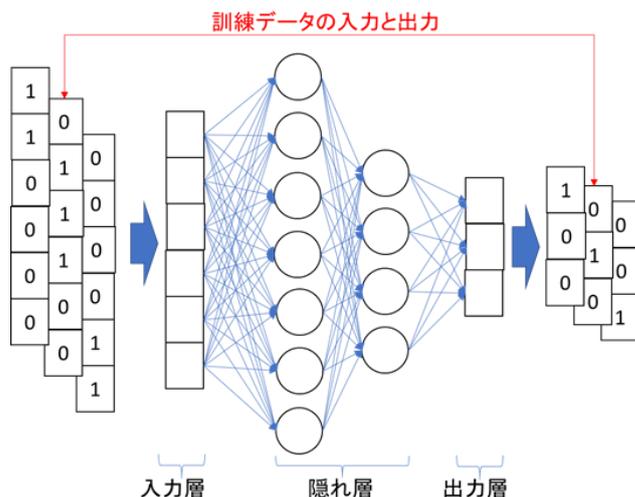


図 2. 訓練課程

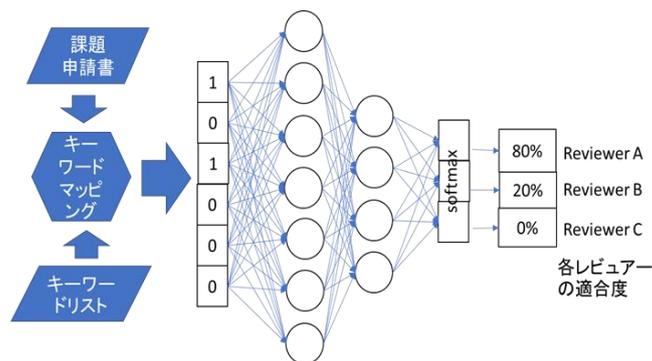


図 3. 実行課程

4 評価計画

4.1 評価の手順

過去の実際の応募課題とレビュアーのデータを用いて、AI と人間がレビュアーの候補選定を行い、人間の判断を正解とみなして AI の判断の一致率を測定し、その違いを分析する。レビュアーの候補選定作業は相互に影響を受けない分離体制を作って行う。すなわち A チームが人間、B チームが AI による作業を行う。手順は以下のとおりである。

- ① A/B チームが協力して、各レビュアー候補が専門とする分野のキーワードを抽出する。科学研究費助成事業データベース[3]、研究者データベース(日本の研究.com) [4]、HPCI ポータルサイトの利用報告書データベース[5]等を参考にする。
- ② A チームはレビュアー候補と課題の双方を科

研費の近い分野の小区分をいくつか束ねた区分に分類する。その結果、各課題に対応するレビュアー候補の集合が得られる。

③ Bチームは①のキーワードを用いてAIを訓練し課題内容からレビュアー候補を選定する。参考として、課題毎にキーワード一致数の多い順にレビュアーを割り当てる方法（キーワード一致法）も実施する。（図4）

④A/Bチームが協力し②と③の結果を比較してAIの出力結果を評価する。

AチームとBチームの一致率(A:B一致率)を、Bチームが推薦したレビュアー上位15人の中にAチームの定義したその課題分野のレビュアー(max15人で分野毎に違う)が何人入っているかの率と定義した。

課題申請書	1	0	1	0	0	0	キーワード一致数	レビュアー割当順位
Reviewer A	1	1	1	0	0	0	2	1
Reviewer B	0	1	1	1	0	0	1	2
Reviewer C	0	0	0	0	1	1	0	3

図4. キーワード一致法によるレビュアー割当

4.2 利用したデータ

本評価において利用したレビュアーおよび応募課題のデータは以下のとおりである。

①レビュアーデータ

ある年度の課題審査においてレビュアー候補として登録されたレビュアーは110名。抽出されたキーワード数は、合計1751で、ユニーク数1026、平均15.9、最大34、最小8となった。キーワードの分布を図5に示す。

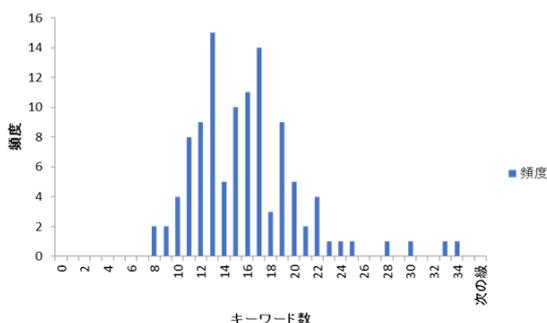


図5. レビューアーのキーワード数分布

②課題データ

上記年度の定期募集に一般枠・若手人材育成枠で応募した98課題を対象とした。

5 MLPの訓練とその分析

5.1 訓練の実施

入力に対する出力の組み合わせ110組（レビュアー毎に1組）のフルデータセットによる訓練を1回とし、入力の順番をランダムに変更して1000回と10000回の訓練を実施した。その結果、図6に示す様に分類の正解率は約360回で100%に達した。ここで「正解率」とは110の入力に対する出力結果が正しくなる割合である。また訓練後のレビュアー識別率は1000回訓練では63.6%となり、10000回訓練で99.2%に上昇した。ここで「レビュアー識別率」とは、あるキーワードパターン（この場合はレビュアーリストの先頭の一人）を入力した時に正解となるレビュアーの確からしさとして出力される値（softmax関数値）である。

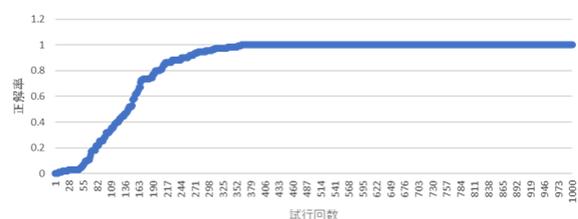


図6. 訓練の状況（1000回）

5.2 訓練結果の分析

上記の訓練を実施したMLPで課題申請書を分類した結果から以下のことが分かる。

①訓練回数を1000回から10000回に増やすとA:B一致率は、37%から47%に向上し、明らかな改善効果があった。

②分野個別に見ても、訓練回数を増やすと課題の分野と一致する専門分野のレビュアーが増加することが確認されたが、誤りも多く実用的なレベルではなかった。ここで「誤り」とは、課題内容から見て妥当性の全くないレビュアーが推薦されたこと、人間が判断したことを言う。

③いくつかのモデルを作成して比較すると、訓練回数と同じであれば、レビュアー識別率は変わらないが、レビュアーの推薦リストは大きく変動することが分かった。しかしA:B一致率の大きな改善は見られなかった。

④AIの結果は、キーワード一致法の結果と同

じとはならない。特に顕著な現象として AI の場合、キーワード一致数が 0 のレビュアーが上位に推薦される場合があった。その理由としては、レビュアーのキーワードパターンと課題のキーワードパターンが似ているほど上位に、違うほど下位にランク付けされ、キーワード一致数 0 の場合には、その中間に取り残されている可能性が考えられるので、今後調査する予定である。

⑤キーワード一致法による A:B 一致率は AI 法における 10000 回の訓練よりも高い値となったが、明らかな誤りも含まれていた。

6 方法の改善

上記の分析結果を踏まえて、次の様な方法の改善を実施した。

6.1 キーワードのチューニング

キーワード一致法においても課題と同じ専門分野のレビュアーが推薦順位の上位に来ないことがあることは、キーワード数が不足していることを示している。そこでレビュアー毎のキーワードを見直して数を増やした。またこのチューニングを、精度を高めるため分野毎の専門家に依頼した。

キーワードチューニングの結果、各レビュアー当たりのキーワードは 20-30 に統一された。全レビュアーのキーワード合計数は 3042、その内重複を除いたキーワード数は 1547 となった。

6.2 訓練データの拡張

訓練する毎に生成されるモデルの挙動が変化するのは、MLP のパラメータの数に対して訓練データが不足しているため、バッチ勾配降下法で訓練されるモデルが一つに収束しないことが原因であると考えられる。そこで実効的に訓練データを拡張する方法として、サンプリングデータ拡張法を考案した。詳細は、付録「サンプリングデータ拡張法」参照。

この方法と「5 MLP の訓練とその分析」と同じデータを用いた訓練の状況を図 7 に示す。(Batch accuracy, Val accuracy の定義は、付録「サンプリングデータ拡張法」参照。) 最終的なレビュアー識別率 (10000 回訓練) は 99.2% から 99.99% に上昇し、A:B 一致率は 53% に改善された。

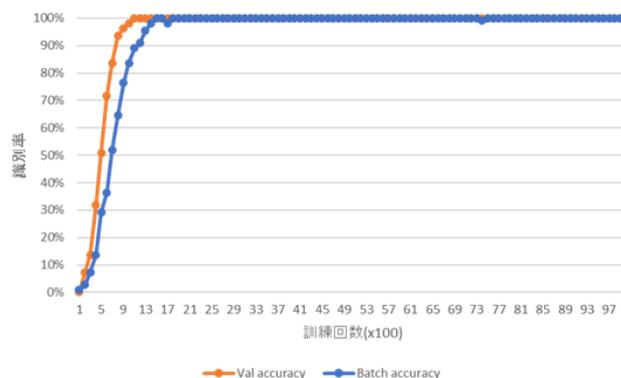


図 7. サンプリングデータ拡張法を用いた訓練

7 改善効果

「6 方法の改善」による効果を、方法毎に A:B 一致率で表 1. に示す。

方法	キーワード チューニング前	キーワード チューニング後
キーワード一致法(KW)	60%	68%
通常方式のAI(AI1)	47%	53%
サンプリングデータ拡張法のAI(AI2)	53%	66%
AI2とKWの組み合わせ	70%	81%

表 1. 各方法による A:B 一致率

A:B 一致率で見ると、キーワードのチューニングにより、いずれの方法においても明らかな改善効果があったと言える。サンプリングデータ拡張法の AI は、ほぼキーワード一致法による A:B 一致率と同等となった。

両方法で選定されたレビュアーを分析した結果、AI で A:B 一致率が低かった研究分野の課題で、キーワード一致法が高い A:B 一致率を示す傾向があることが分かった。そこで、サンプリングデータ拡張法の AI でレビュアーの適合度を評価した後に、キーワード一致数の最高位レビュアーの専門分野にバイアスを掛ける方法を試した (AI2 と KW の組み合わせ)。その結果、人間の評価の 8 割を再現できることが分かった (表 1)。残りの 2 割は、キーワード一致数 1 位のレビュアーの専門分野が課題の専門分野と相違していることを示す。

課題とレビュアーの専門分野を排他的にかつ唯一に定義できれば、キーワードをチューニングしてさらに一致率を向上させることは可能であると

考えられる。しかしこの方法では、基本的に単一分野のレビュアーを選定することになるので、分野融合的な課題に対応できない。今後は A:B 一致率という評価基準を再検討する必要があると考えられる。

8 今後の課題と可能性

AI によるレビュアー選定の精度を高めていくために、以下のような改善策を検討していく予定である。

8.1 ニューラルネットワークのハイパーパラメータの最適化

ニューラルネットワークの一種である MLP には、操作可能なパラメータとして、隠れ層の数、隠れ層当たりのニューロン数、活性化関数、サンプリングデータ拡張法のサンプル数、等があり、広い範囲でチューニングの余地が残されている。

8.2 他の機械学習アルゴリズムの採用

今回用いた単純な MLP の他にも、深層ニューラルネットワークのアルゴリズムは多数存在するので、より有効なものを発見できる可能性はある。

8.3 レビュアーのキーワード作成の自動化

本方式ではレビュアーを追加する度に専門分野を定義するキーワードを人間が作成する必要がある。そこでインターネット上の情報からレビュアーのキーワードを自動的に抽出するツールを開発することにより、省力化やキーワード数の増加等の改善が考えられる。

8.4 より詳細な課題情報の取り込み

web 上に登録された課題概要だけではなく、追加資料となっている課題申請書に書かれた情報を利用できれば、より詳細な課題情報を取り込むことが可能となる。

8.5 新しい評価基準の検討

現在の評価基準である A:B 一致率は、人間が課題毎に決めた単一の研究分野に基づいている。新しく開拓された研究領域に対応する分類方法や、分野融合的な課題を正しく評価する方法を検討する必要がある。

8.6 キーワードの正規化

レビュアー毎のキーワード作成において、作成する語数や、複合語、一般用語、同意語、英語

やカタカナ等表記の違いの扱いをどうルール化すべきかについては今後の課題となっている。また課題申請およびレビュアー登録のシステムを機能強化して、キーワードについては正規化された用語リストから選ぶ様にすることも考えられる。その際には、分野、手法、モデル、対象とする現象等キーワードを構造化して整理する工夫も検討する必要がある。

8.7 キーワードマッチングの限界からの脱却

キーワードを使う限り、言葉の表層的な一致性のみを使うことから来る限界があるかもしれない。レビュアーの知見と課題の内容をキーワードではなく、意味空間に写像して相互の距離を比較できれば、より良い判断ができる可能性がある。テキストマイニング等の手法を利用して課題申請書やレビュアーの著書を解析することが考えられる。

9 まとめ

ここで評価に用いた単純な MLP においても、キーワード一致法と組み合わせることにより、人間の判断との一致率を 81%まで高めることができた。しかしこの方法では、分野融合的な課題に対応できないという問題を引き起こす可能性があるため、必ずしも良い解とはならない。

また AI 法単独で人間の判断との一致率は 66%となるが、キーワード一致法に対して明確な優位性が見えず、改善の余地が大きいことが判明した。現状では、AI によってレビュアーの選定を完全に自動化することはできないが、人間による選定の候補者案を作成することは可能であり、特に分野融合的な研究においては有効であると考えられる。

今回利用したのは、MLP による分類問題であるので、課題申請書の特徴からレビュアーを識別させること、すなわち正解のない問題（この課題申請書を書いたのはレビュアー本人でないという意味で）を解かせること自体に限界があるかもしれない。

今後は、その限界を考慮しつつうまく活用する方法を、実際の課題選定に応用しながら模索して行きたいと考える。

謝辞

本論文の作成にあたり的確な助言を頂いた一般財団法人高度情報科学技術研究機構の皆様へ感謝します。また特に、分野専門家としてキーワードチューニングに協力して頂いた、太田幸宏、岡田達夫、須永泰弘、澤井秀朋、山岸孝輝の各氏へ感謝します。

付録：サンプリングデータ拡張法

レビュアーのキーワードリストから、ランダムサンプリングによって訓練データの再生成を繰り返し、ミニバッチとして、指定した回数だけ逐次的にモデルの改良を行う（今回は勾配降下法を用いている）方法として以下の手法を用いた。

改善前の方法では、レビュアー (i) 毎に、専門分野を現すキーワードの集合 (K_i) を定義し、入力 K_i に対応する出力が i となるように訓練を行う。従って訓練データは、基本的にレビュアーの数 (N) 対の 1 セットしかなく、そのままでは訓練データが十分確保できない。

そこで、次のように訓練データを再生成しながら機械学習を行うアルゴリズムを開発した。

- ① ミニバッチ (j 回目) 毎に②~④を繰り返す。
- ② レビューアー (i) のキーワード集合 K_i からランダムサンプリングにより一定数 (今回は 8) の語を抽出して入力 K_{ij} とし、出力を i とするデータ対を生成する。(実際の実装では、入力データはワンホットエンコーディングであり、 K_{ij} に対応する位置が 1、他が 0 となる一次元ベクトルが生成される。)
- ③ 上記のデータ対作成をレビュアーの数 (N) だけ繰り返し (データ拡張に相当)、ミニバッチを構成する一組の訓練データ $\{K_{ij}, i\}$ として訓練オペレーションを実行する。
- ④ 訓練されたモデルの評価のため、 $\{K_{ij}, i\}$ を用いた時の正解率 (Batch accuracy) と、フルキーワードデータ $\{K_i, i\}$ を用いた時の正解率 (Val accuracy) を測定する。

またここで指定するサンプリング数は、レビュアーのキーワードから同時にいくつを訓練の入力データとして抽出するかを指定するものである。訓練されたモデルは、実際には課題申請書の中で一致するキーワードを入力として、レビュアーの

適合度を出力するものである。従って、サンプリング数の指定においては、そのキーワードが「課題申請を書く時に、いくつ同時に使われるか」を考慮することが望ましいと考えられる。

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
- [2] Aurelien Geron (著), 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
- [3] 科学研究費助成事業データベース, <https://kaken.nii.ac.jp/ja/>
- [4] 研究者データベース (日本の研究.com), <https://research-er.jp/about/database>
- [5] 利用報告書データベース (HPCI ポータルサイト) , <https://www.hpci-office.jp/annex/usrrep/>