



公開重みで、 歴代首位級のスコアへ

Kaggle Predicting Molecular Properties

統合戦略・EDA・AI協働で進めたRTX3060再現研究

<https://www.kaggle.com/competitions/champs-scalar-coupling>

菜園の柱 / 2026.10.27



今回の結論：公開重みの再生で、歴代首位級へ

Private -3.22754

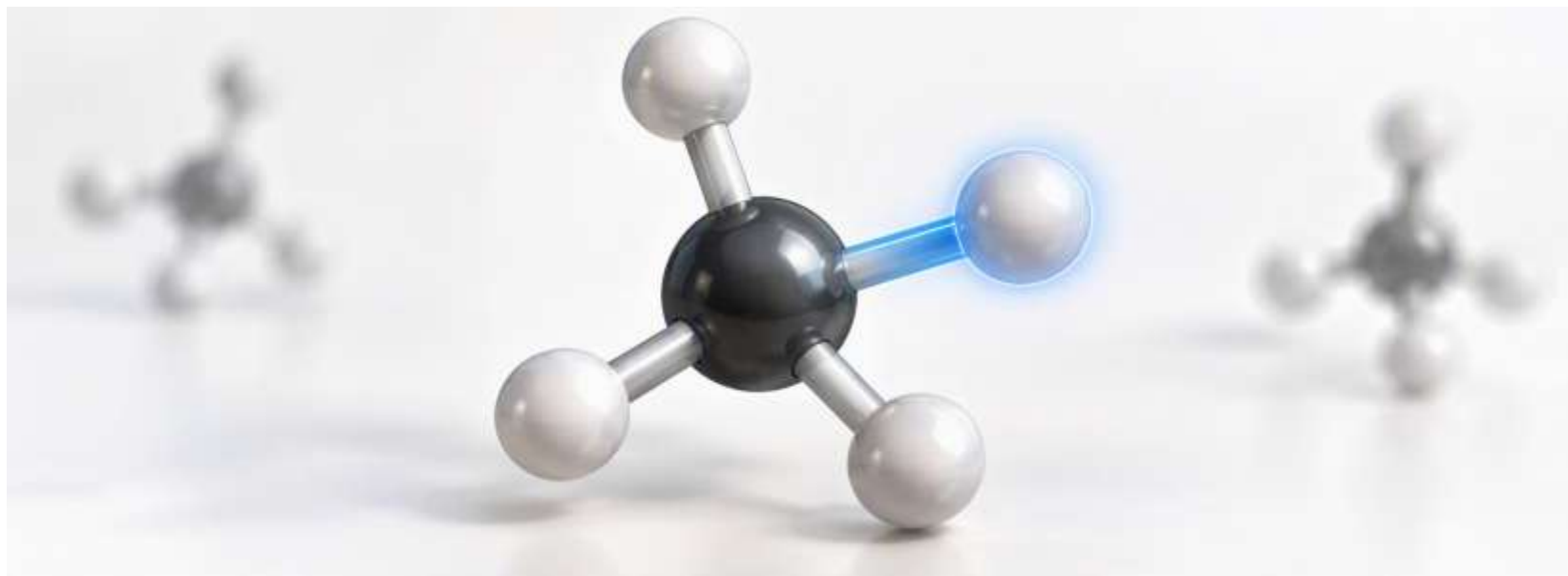
RTX3060（12GB）1台で、優勝解法の公開13モデルを再生

採用した戦略	今回の到達点
第1段階：公開資産を再現	推論・公式集約・検算を経て、首位との差 +0.01214（今回は高い＝悪い）
第2段階：新技術で上積み	モデル内の改善は確認。追加採用の根拠には届かず。

Private＝非公開評価データのスコア（低いほど良い）。評価の定義はp.3、詳しい比較はp.21。
公開13本は再学習せず推論。自前学習の検証実験は別に実施した。

どんなコンペか？

入力：元素と3次元座標、対象の原子対



例：H [1] と C [0] の $J = 84.8076$ Hz

評価：8タイプそれぞれの平均絶対誤差（MAE）の自然対数を平均。低いほど良い。
スコア-3.2なら、タイプ別MAEの幾何平均が約0.04 Hz（全行MAEとは別）。

J は指定した2原子に1値。 J も、対数を使う評価スコアも、負の値をとる。

3サービスの調査を、一つの実行戦略へ

共通の条件

3つの調査案

比較・精査して一本化

**共通プロンプト**CHAMPS
RTX3060 12GB
二段階目標**ChatGPT**旧データ・評価固定 / 忠実再現
TensorNet・MACE / OOF統合**Claude.ai**歴史的成績・優勝コード / 異種ensemble
三角更新 / 計算予算**Gemini**原子対テンソル縮約
電子状態の補助学習 / pruning等**Codexで比較・統合****Claudeでチェック・修正**

最終成果物

ChatGPT-Total-12GB-v3.1.md

第1段階

2019優勝解法の最高スコア再現

第2段階

2026年時点の技術でさらなる向上

どんなプロンプトで、どう戦略を作ったか

共通プロンプトの要約：目的・制約・成果物まで指定する

Deep Research系でCHAMPSを調査し、①2019年優勝級の解法を再現、②新技術で上積みを狙う。モデル紹介だけでなく、公開コード・評価方法・計算資源・実装手順まで示す。歴史的実績と未検証案、実測値と推定値を区別し、不明な情報は創作しない。

指定したこと	プロンプトに含めた具体条件
データ・検証	旧データを統一。分子単位CV。test正解・代理ラベルの利用を禁止。
比較・実行計画	上位解法と公開コード、主戦略と代替案、計算コスト・優先順位を提示。

ChatGPT案 → Claude確認 → 3サービス調査 → Codex統合 → Claude確認 → 実行

同じ依頼文で3案を比較。サービスごとの調査モードは完全同一ではない。

Codex採用理由：ローカル実行から、許可範囲内のKaggle提出・採点取得まで連続実行。

原文：common-prompt-body.md / 統合戦略：ChatGPT-Total-12GB-v3.1.md

二段階の目標：公開解法を再生し、新技術で上積みを狙う

公開重みはKaggleで採点。新技術は自前の検証分子で比較し、評価を混同しない。

STAGE 01 / REPRODUCE

目標：公開重みで優勝級スコアへ

STAGE 02 / IMPROVE

2026年時点の技術でさらなる向上



優勝チーム hybrid | Bosch Research

同システムの学習済み13本を使用（次ページ）
タイプ別9本 → 中央5本を平均

完了：13本の推論・公式集約・採点

TensorNet-small | 原子対方向読み出し

方向読み出し・幅・学習時間を検証
さらなる向上は目標（未達成）

未実施：MACE-small・タイプ別OOFブレンド

旧データで再生。自前モデルは固定分子foldで評価

第1段階で使うA～M：Graph Transformer系の13モデル

13種類の別アルゴリズムや、13分割のfoldではない。構成・重みが異なる13本。

公式名	幅（内部表現の次元）	層数	特徴
A・B・C	650	10	同じ主要設定、別々の学習済み重み
D・E・F	600～800	10～11	幅・内部層などの構成が異なる
G・H・I・J・K	600～700	11～16	深さ・幅・出力部などが異なる
L	750	18	四つ組の情報も使う構成
M	660	11	別構成の学習済みモデル

公開A=この13本のうち1本。13本を推論 → タイプ別9本 → 中央5本平均。

勝者の学習環境：Linux13台 / RTX2080 Ti×20 + V100×8（28GPU）。

今回：RTX3060 12GB×1台で、保存された重みを使って推論した。

Codexを司令塔に、研究と資料制作を並行する

今回の制作フロー：判断・研究・資料制作を分担する

01 APPROVAL

人間

目的
優先順位
見た目の承認



02 COMMAND

Codex

研究実装・根拠確認
制作指示・結果検査

Codex内蔵ブラウザ

ブラウザUI操作
API連携ではない
指示・素材添付



出力・プレビュー確認



03 PRODUCTION

Copilot

共通デザイン
概念イラスト
PPTX制作



PARALLEL

ローカルGPU：学習・checkpoint保存

Codexの利用枠とは別に実行継続

研究の判断・復旧枠を残し、学習待ち時間に資料制作

Codexの利用上限：5時間枠を研究と資料作成に配分

残り枠の目安	運用
40%	資料作成の警告。次の作業を小さく区切る
30%以下	非緊急の資料作成を止め、研究の判断・復旧枠を残す
15%以下	研究の必須対応に限定
自然回復後	残量を確認し、GPU計算中に資料作業を再開

夜間学習はユーザー操作不要にする。週枠も確認し、発表期限までの必要作業を優先。

研究期間は約70時間、主要計算は45時間19分

目的と位置づけ	時間	公開Aとの対応・使い道
第1段階の主成果：公開重み再生	推論5時間53分	A～Mの学習済み重みをそのまま使用
第1段階関連：比較基準の自前学習	学習19時間21分	公開Aと同じ構造・公開重み未使用。 学習条件は一部変更して一から学習
第2段階：現代モデルの追加探索	学習20時間05分	fold0：自前2019系と比較 fold1：TensorNet同士を比較

公開重みは元train全体を学習済み。未学習の検証分子を残すため、自前2019系を別に学習。

実行では比較用の自前学習が先行し、公開13本の採点は10/26午後。
別学習39時間26分／主要計算45時間19分≒87%。再生には不要で、最終提出に未使用。

研究結果まで69時間49分（10/23 22:41 → 10/26 20:30）。睡眠・外出・回復待ちを含む。
計算時間とは合算しない。人の実操作時間は未計測。87%は主要計算記録内の割合。

ここからデータ理解：CSVの読み方

EDAは公開重みの再生とは独立。構造の仮説と、誤差を群別に見る軸を作る。

train.csv：1行＝原子対と正解J

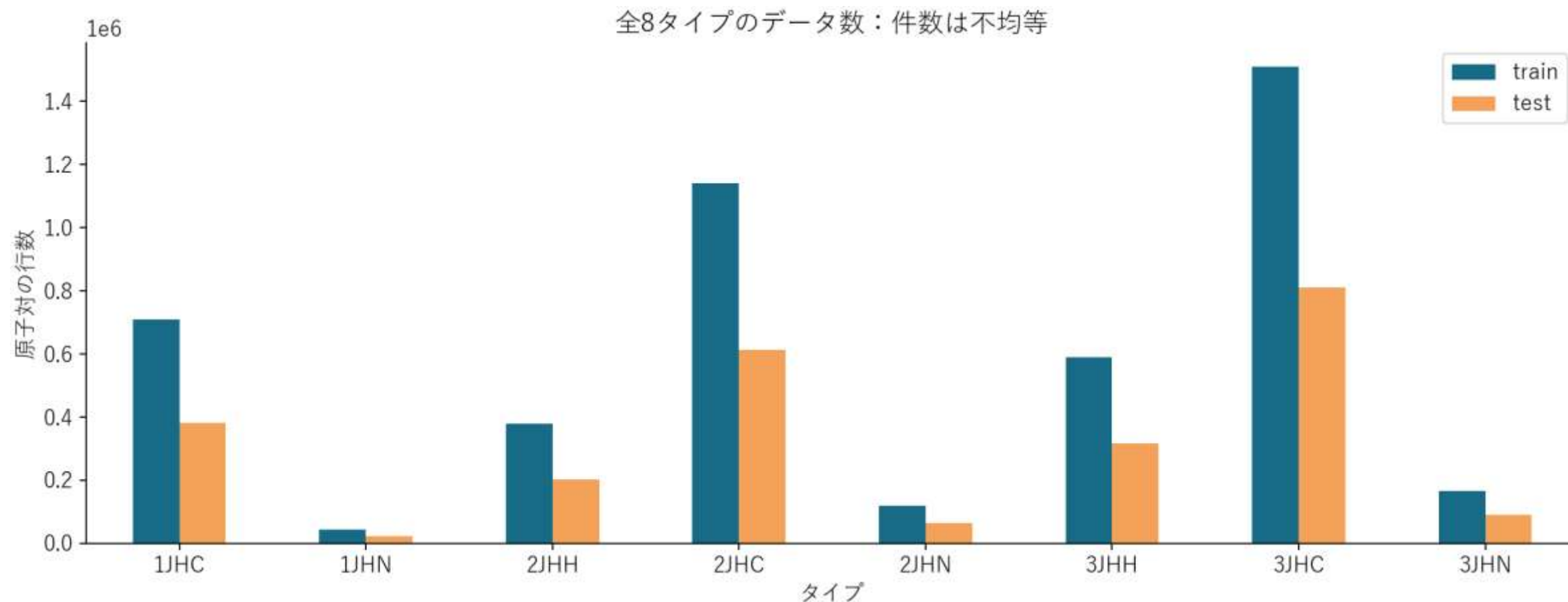
id	molecule_name	index 0	index 1	type	J [Hz]
0	dsgdb9nsd_000001	1	0	1JHC	84.8076
1	dsgdb9nsd_000001	1	2	2JHH	-11.257
2	dsgdb9nsd_000001	1	3	2JHH	-11.2548

structures.csv：同じ分子 dsgdb9nsd_000001 の5原子

atom_index	atom	x [Å]	y [Å]	z [Å]
0	C	-0.01270	1.08580	0.00800
1	H	0.00215	-0.00603	0.00198
2	H	1.01173	1.46375	0.00028
3	H	-0.54082	1.44753	-0.87664
4	H	-0.52381	1.43793	0.90640

旧データ＝2019年コンペ時のtrain/test。列名は略記。分子名×原子番号で照合（1 Å＝10⁻¹⁰ m）。

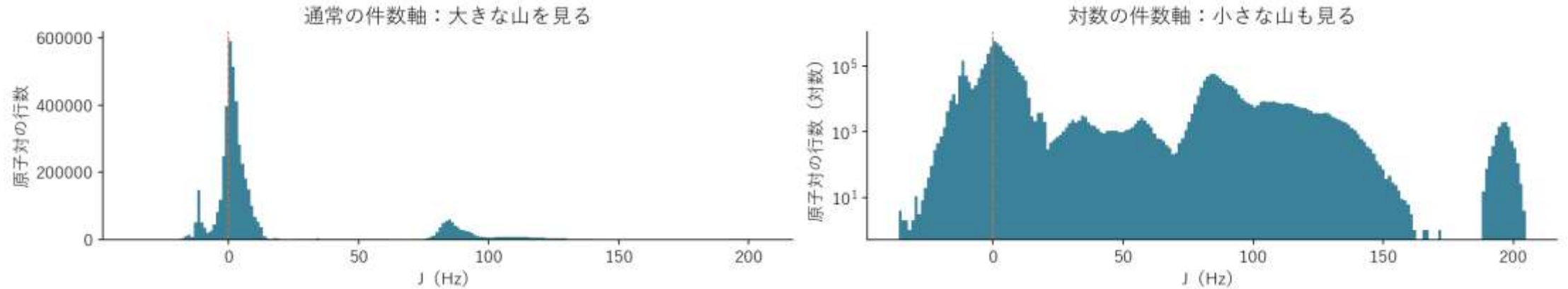
予測対象は全8タイプ：件数は大きく異なる



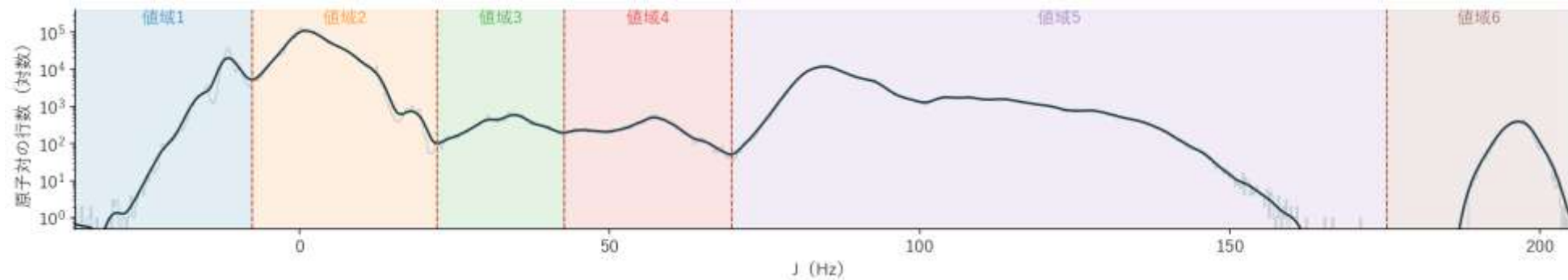
数字＝隔たる結合の数、後ろの文字＝原子の組。例：1JHCは結合1本で隔たるH-C。
全8種：1JHC / 1JHN / 2JHH / 2JHC / 2JHN / 3JHH / 3JHC / 3JHN（図と同順）。

Jの全体分布：通常・対数表示と、谷からの6値域

旧train全4,658,147原子対。上段は同じ200ビンで、右だけ件数軸を対数表示。

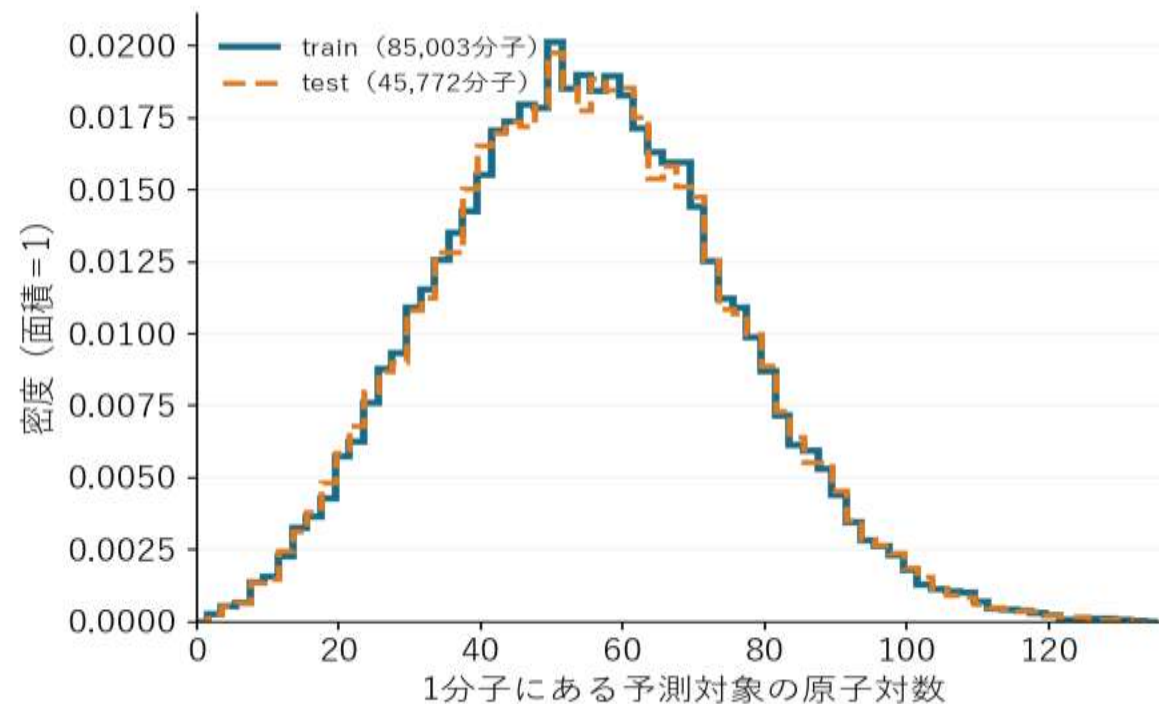
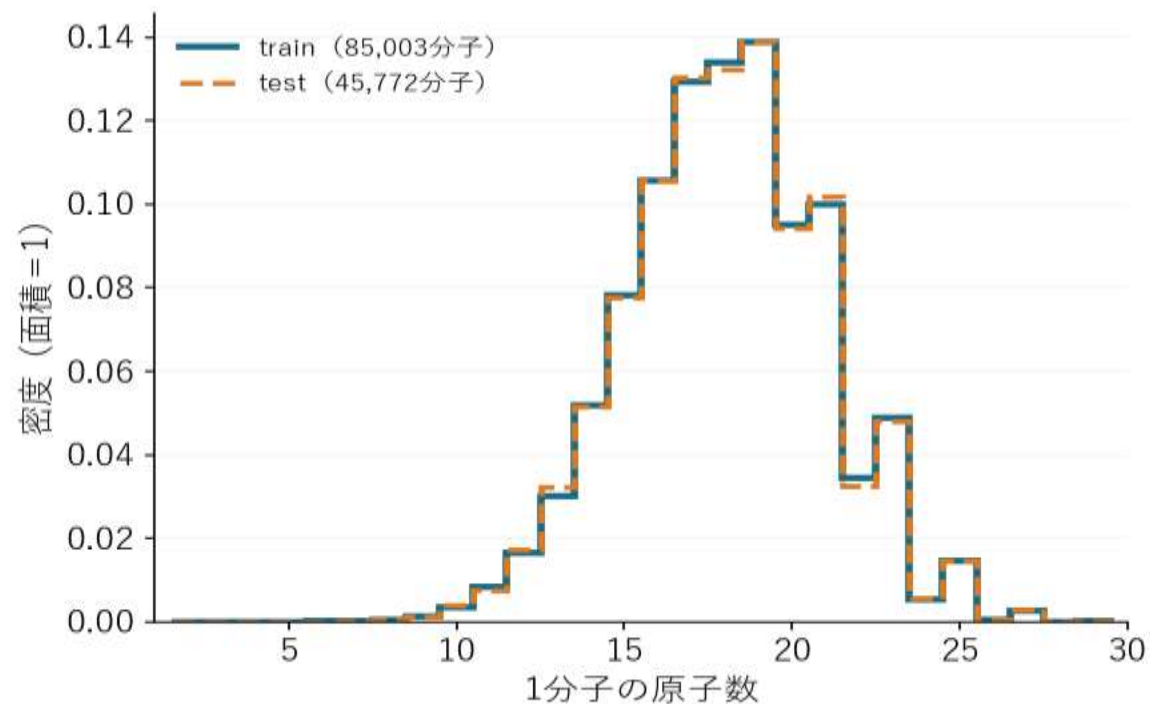


小さな山・谷を平滑化して6値域を観察し、構造のk=6という探索案へ



下段も件数軸は対数。0.25Hzビン・帯域幅1Hzで平滑化。J自体は対数変換していない。
Jの値域と分子の構造クラスは別分類。一対一対応や最適なkを証明したものではない。

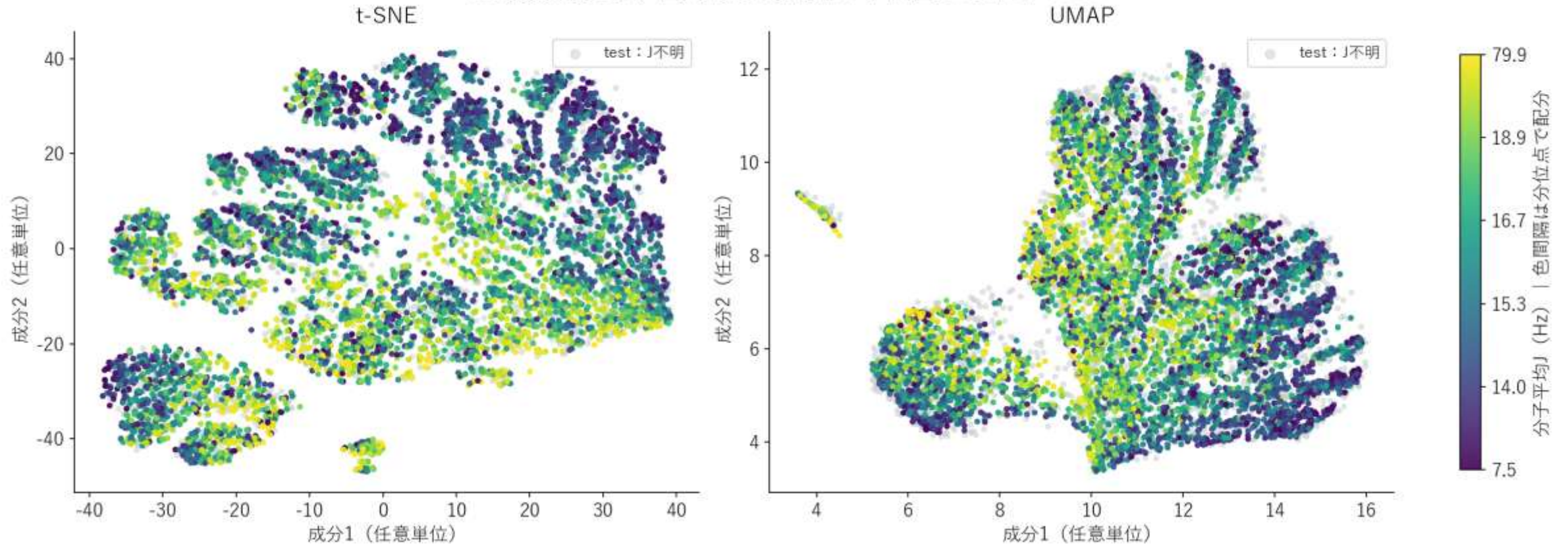
分子サイズと、原子対数の分布



青実線=train、橙破線=test。表示した分子サイズ・原子対数の分布はほぼ同じ。

Jで色分けすると、地図上に偏りが見える

同じ分子座標・同じJ：分位点に合わせて色を配分
Jを対数変換せず、色間隔だけ非線形に。灰色はJが未知のtest



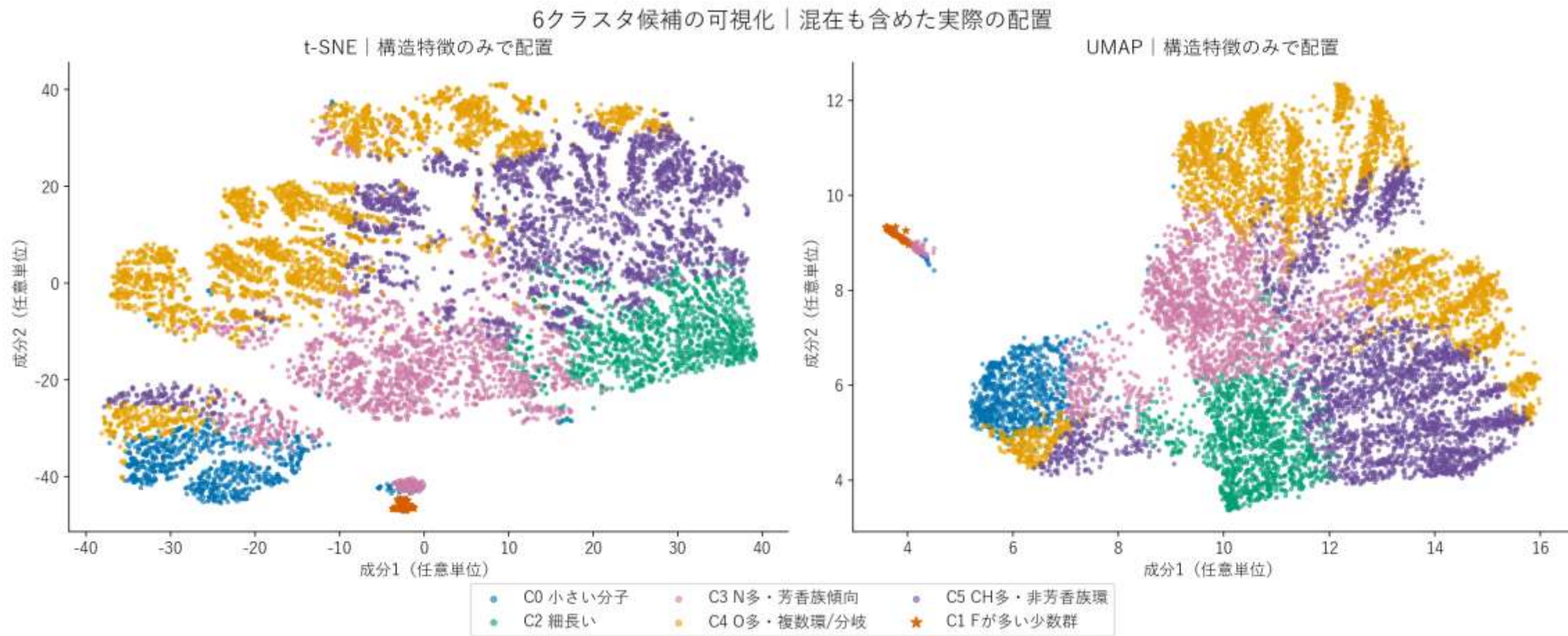
点＝分子。15構造特徴から作った2次元地図（t-SNE・UMAP）。分子平均Jの分位点で色分け。
色間隔は非線形。元のJ・座標は不変。灰色はJ未知のtest。タイプ構成も平均Jに影響する。

Jの谷をヒントにk=6：構造の特徴を言葉にする

群	意味づけの仮説
C0	小さい分子
C1	Fが多い少数群
C2	細長く、複数環や分岐が比較的少ない
C3	Nが多く、芳香族の傾向
C4	Oが比較的多く、複数環・分岐の傾向
C5	CHが多く、非芳香族環の傾向

Jは入力せず、15構造特徴だけでk-means。群の解釈を仮説として記述し、次の地図で確認。

6クラスタ案をt-SNE・UMAP上で確認する



表示条件は見やすさで選択。島の分離は、真の化学分類やスコア改善の証拠ではない。

クラスタ距離の追加は、今回のOOFで改善せず

教材用LightGBMの全5fold・学習外予測（OOF）で比較。低いほど良い。

① 基本特徴を順に追加	5fold OOFスコア
原子対の距離だけ	+0.716809
+ 周辺の幾何情報	-0.163717
+ 元素組成	-0.179634
+ 分子全体の形状【基準】	-0.180087

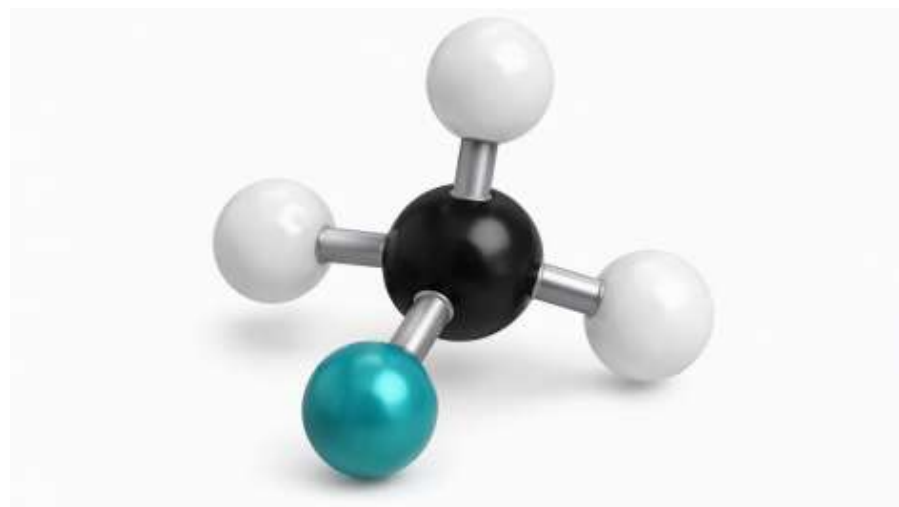
② ①の最後の行を基準に比較	OOFスコア	差（追加-基準）
基準 + 6クラスタ中心への距離	-0.179598	+0.000489（悪化）

不採用は②の「クラスタ中心への距離特徴量」。①の幾何・組成等とは区別。

指標は8タイプMAEの対数平均。教材用のモデルと全5fold評価で、p21・22と直接比べない。

第1段階：Graph Transformer：分子の情報を集めてJを予測

分子の形を手がかりに、指定した原子対の数値を推定する



対象の H-C 原子対

① 分子を表現

原子・原子対・三原子組
元素・距離・角度などを入力



② attention

距離なども考慮して、分子内の
情報を重み付きで集める



③ 回帰

対象原子対の種類に応じた
出力から J [Hz] を得る

グラフを扱うニューラルネットワークの仲間。近傍集約GNNとは情報の集め方が異なる
文章生成ではなく、教師ありの数値回帰

第1段階：Graph Transformer 13本を公式ルールで集約

工程	何をしたか
① 構造を特徴へ	元素・座標から原子、全原子対、三原子組の表現を作る。 距離・角度・化学環境などを入力する。
② 13本で予測	A～Mは構成・学習済み重みの異なる同系統モデル。 旧testの2,505,542原子対についてJを予測する。
③ タイプ別に9本	作者が定めたタイプ別のモデル選択表（mask）をそのまま使う。 今回のKaggle順位表（LB）を見て組合せを調整していない。
④ 中央5本を平均	各原子対の9予測を並べ、上下2本ずつを除く。 残る中央5本の平均を提出値にする。

公開モデルAは13本のうちの1本。公開重みは作者が全trainで学習済み。

配布元：Bosch Research公式GitHub（解法を再現するためのコード・設定・取得手順）
github.com/boschresearch/BCAI_kaggle_CHAMPS ／ [get_saved_models.sh](#) → Google Drive

公開重み13本で、歴代首位級のスコアへ

Private -3.22754

Core i7-11700K ・ RAM128GB + RTX3060 12GB ×1台

比較対象	Private（低いほど良い）
公開A・単体の再生	-3.00370
今回：公開A～Mの13本集約	-3.22754
歴史首位：優勝チーム hybrid	-3.23968

歴史首位より0.01214高い（悪い）。公開資産の再生が、今回の主成果。

作者資料は最終提出を13モデル集約と報告。今回の公開13本再生には差が残った。
歴史スコアとの比較で、当時の順位取得ではない。勝者の学習コストとは別。

第2段階：内部比較は改善、主成果への追加は未達

数値は検証スコア（低いほど良い）。方向読み出しの仕組みは付録p.27。

TensorNetの条件	fold0検証	fold1検証
初期64・20epoch	-1.164366	-1.180464
方向64・20epoch	-1.493998	-1.437380
方向128・20epoch	-1.672117	-1.712256
方向128・100+20epoch：通常lr	-2.003803	-1.992448
同：追加20epochだけ低lr	-2.208615	-2.212999
比較：自前2019系・20+3epoch	-2.368256	未実施

低lrは101～120epochだけ（0.0005→0.0001）。100epochまでは同じ親重み。

差（低lr-通常lr）：fold0 -0.204811 / fold1 -0.220551。

自前2019系＝公開Aと同じ構造。公開重み未使用、固定fold・BF16・最後の3epochは低lr。
TensorNetは内部で改善したが、fold0の自前2019系には未達。学習条件は同一ではない。

指標は8タイプMAEの対数平均。融合はOOF不足で未検証。120epoch重みのtest推論・提出も未実施。

到達点と未達の考察：分かったこと、残る検証

論点	確認できたこと／解釈の範囲
公開13本と首位の差	Private -3.22754、首位との差 +0.01214。 原本集約は再検算済み。差の原因は未特定。
RTX3060の影響	計算資源は学習規模・探索時間を制約する。 同じ公開重みの推論が必ず低精度になる根拠はない。
現代モデルが未達	小型2層TensorNetを原子対回帰へ適応。 幅・時間・学習率で改善したが、fold0の自前2019系を超えず。

戦略の評価：公開資産再生は成功。追加探索の時間配分には反省が残る。

目標は再生→上積み。実行では自前学習が先行し、13本の採点は10/26午後。
首位との差は計算精度・前処理等の影響が候補。RTX3060固有の精度不足とは未確認。

EDA距離特徴は評価で改善せず不採用。期限で追加探索を終了（p.18）。
今後の混合は誤差の補完性を検証。公開重みの学習内予測では、公正な評価ができない。

Summary : 公開資産の再現を主成果に、次の課題を残す

今回得たもの	知見と、次に検証すること
成果	RTX3060 1台で公開13本を再生し、Private -3.22754。 歴史首位より0.01214高い（悪い）ところまで到達。
戦略と進め方	公開資産の選択は成功。最終提出に未使用の自前学習は主要計算の約87%。 再生を先に確保し、追加探索の予算・打切り条件を決めるべきだった。
EDAと第2段階	構造理解・仮説生成と、スコア改善の実証は分ける。 6クラスタ距離の入力は不採用。現代モデルは内部改善にとどまり、上積み未達。
残る課題	首位との差の原因を切り分ける。現代モデルの構成・最適化を検証。 各モデルが未学習の分子への予測をそろえ、混合の効果を公正に評価する。

コード・固定分子fold・重みの来歴・実測時間を残し、再開と再現を可能にした。

本編はここまで。付録：評価方法とTensorNetの仕組み。

付録

APPENDIX

評価・検証の定義
第2段階のモデル説明



旧データと分子単位の固定検証

データ	原子対の行数	分子数
旧train	4,658,147	85,003
旧test	2,505,542	45,772

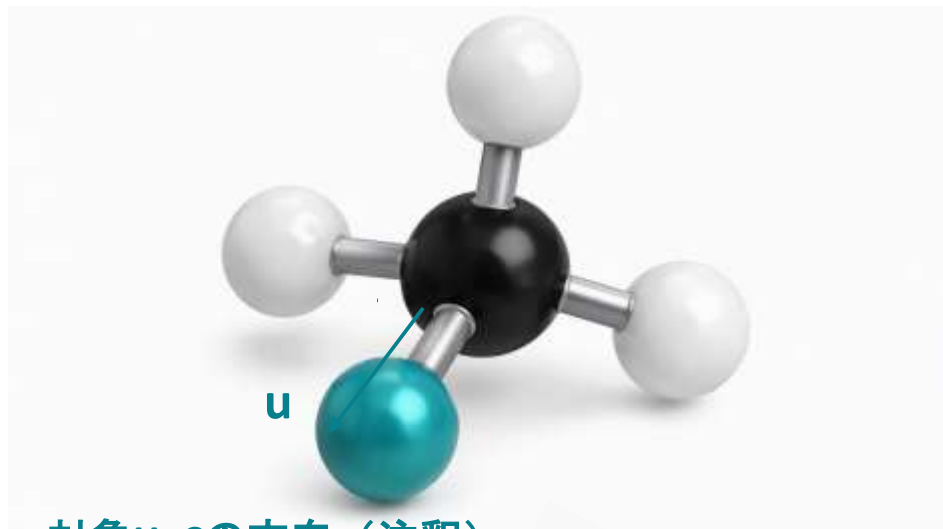
同じ分子の原子対が学習側と検証側にまたがらないよう分割。

評価：8タイプごとのMAEを計算し、自然対数を平均。
低いほど良い。MAEが1 Hz未満なら対数は負になる。

前処理は学習分子だけでfit。Kaggle、単一fold、完全OOFの評価を区別する。

TensorNet : 原子対の方向から情報を読み出す

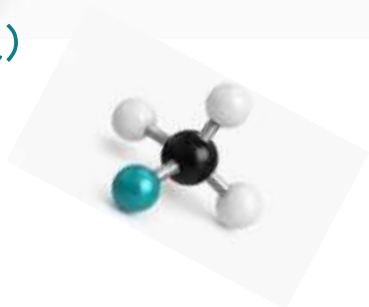
分子の向きが変わっても、予測するJは同じ



対象H-Cの方向 (注釈)

分子全体を回転

内部配置は変えない



① 幾何情報を交換

元素・3次元座標から近傍原子と情報交換し
原子ごとの方向表現を作る

② 対象原子対に沿って読む

両端の3×3テンソルを
方向uに沿ってまとめる ($u^T X u$)

③ Jを回帰

両端の特徴・距離・タイプと合わせ
J [Hz] を出力

今回 : TensorNet 2層、幅64/128。公式バックボーン+独自の原子対読み出し
方向表現は回転に対応して変わり、最終Jは回転・平行移動に依存しない