WILEY

検証試験: Wiley Database of Predicted IR Spectra

要約

Wiley Database of Predicted IR Spectra は、Wiley Science Solutions がリリースした最新のスペクトルライブラリです。予測スペクトルの精度と妥当性を確認するために、外部検証（external validation）試験のフレームワークを用いて検証テストを開発し、実施しました。本試験は、最終的な予測を現実世界のシナリオで評価することで構成されています。実測のターゲットに対して検索された実測の重複レコード（注：同一化合物を異なる装置で実際に測定されたものなどで、スペクトル形状は完全に一致しないもの）と、予測ターゲットに対して検索された同じ重複レコードセットの二つのヒットリストを分析するテストです。これらのテストの結果、予測データでは 85%超の確率で、正答が上位 10 ヒットに含まれました。一方、実測データでは 94%超の確率で、正答が上位 10 ヒットに含まれました。これらのテストでは、予測データが、重複レコードを用いた実測データでの検索と同等のレベルで、同一のテストを実施することが示されました。

緒言

Wiley Database of Predicted IR Spectra[1] は、適切な実測のスペクトルライブラリの一致スペクトルが得られない場合に、未知の物質を同定するために所定の化学空間内でより多くの化合物で赤外スペクトルを分析できるように設計されています。この予測 IR ライブラリを実測のライブラリと併用すると、未知のスペクトルの組成を同定し、化合物の関連する構造的構成と構成官能基についての洞察が得られるツールとして使用できます。本検証試験の目的は、予測ライブラリが実測のデータベースと同等のレベルで機能するか観察することです。

予測ライブラリの新規性と実験的性質により、外部検証試験を実施することを決定しました[2]。本試験は、現実的なシナリオで予測ライブラリを評価するように設計されました。スペクトル検索ヒットリストの分析は、ユーザーが最終製品のライブラリの使用をどのように体験するか、ということに関する代表的なフィールドテストとして使用されます。これは、テストセットの計算による評価を超えたテストです。

スペクトル検索ソフトウェアとスペクトル参照データベースの併用は、未知のスペクトルの特性解析と同定の根幹となるものです。スペクトル一致度の比較の結果、または参照データベースへの「ヒット」には、スペクトルがデータベースのスペクトルとどの程度一致するかランク順にするためのに、データベース一致度を示すヒット品質インデックス（HQI）スコアが割り当てられます。本試験では、KnowItAll[3] の SearchIt アプリケーションを使用して一連のスペクトル検索を実行し、スペクトルを実測の参照データセットおよび予測参照データセットの両方と比較しました。

方法

データ

Wiley Database of Predicted IR Spectra[1] 用のテストを作成するには、コンピュータモデリングで使用される標準的な訓練/テストデータの分割（train/test split）に加えて、実測の参照データを使用することが最良の方法であると判断しました。Wiley の IR スペクトルデータベースコレクション[4] の膨大なカタログ内で、実測データの重複レコードを検索する際に、ヒットリストの平均値を決定しました。実験データと正確に比較するために、重複レコード検索の比較のベースラインを確立し、検索シナリオで実験データがどのように機能するか（すなわち、重複レコードを考慮すると、実測のスペクトルのライブラリからの再現率が予測ライブラリからの再現率に対してどの程度優れているか）判定しました。これは、結果の CSV を自動的にエクスポートする自動バッチ方式で、KnowItAll[3] ソフトウェアを使用して実施しました。

重複レコードのデータは、テストセットのレコードから取得し、モデルに現実的な「テスト」を提供するために、テストセット内のレコードと全く同じ構造を持つあらゆるレコードを除外しました。また、1 つの構造につき 1 つの予測しかできないため、テストセット内のレコードに重複レコードを含めることはできません。つまり、Wiley のデータカタログに重複レコード分析で使用できる重複レコードが残っているということです。Wiley のデータカタログは、両検証テストに使用された重複レコードの取得元です。テストセットを使用データの 5%とした場合、3712 のモデルテストセットから 33 の重複レコードが使用され、2 回目のテストでテストセットを使用データの 10%に拡大した場合は 7424 のモデルテストセットから 427 の重複レコードが使用されました。

SearchIt

「SearchIt」は、選択した参照データベースに対してスペクトル、ピーク、化学構造、または特性値を検索するために使用される KnowItAll[3] のアプリケーションです。このテストでは、スペクトル検索のみを使用しました。KnowItAll[3] の特許取得済みの最適化補正（スペクトル自動補正）を使用した場合と使用しない場合で、KnowItAll[3] の相関（Correlation）検索アルゴリズムを使用しました。最適化補正は、とりわけスペクトルから不純物を除去します。これは、検索の一致度に影響を与え、多くの場合、改善することができます。

ヒットリスト分析

ヒットリストの出力は、カスタム開発ツールを使用して実行しました。分析には、テストセットから得られた 2 セットの異なる重複レコードを使用しました。これらの派生セットは、ヒットリスト内の構造の完全一致検索によって相手のレコードを見つけ、予測データセット内でも構造が一致するように設計されています。検証テストが正しく機能するには、正確なヒットリストの結果を得るために、各データベース内で一致する化合物は一つだけでなければなりません。実際には、次の三つのデータ

ベースがあります。1) 一つの予測テストセット、2) 実験テストセット、3) 他の二つのセットを検索するためのターゲット重複レコードセット。

自動分析では、予測テストセットと実測のテストセットの両方のスペクトルが、重複レコードターゲットデータベースにて検索されました。最初に、スペクトル検索を使用してヒットリストを生成し、次に構造式での完全一致検索をヒットリストに基づいて実行し、各ヒットリストのフレーム内でターゲット重複レコードの位置を見つけました。これは、実測データに対する予測データ、および実測の重複レコードに対する実測データのベースラインを生成するために行われました。

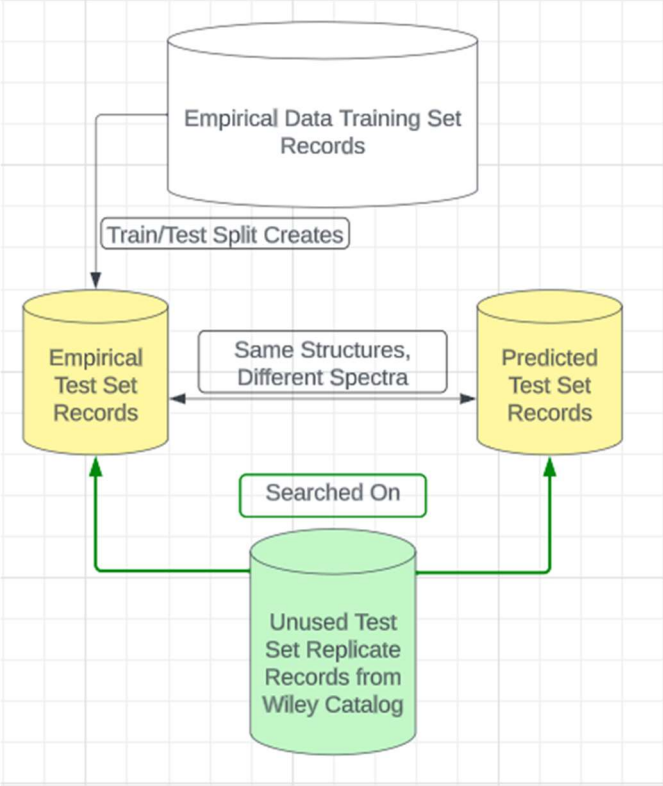

図 1：訓練/テストデータの分割のフローチャートと検証試験の視覚化

結果

1 回目の検証テスト

自動検証の結果は、実測の参照データベースと予測データベース間において、ヒットリスト結果の統計的類似性が高いことを示しています。

実測のテストセットデータベースを実測の重複レコードで検索した場合、94%の確率で上位 10 ヒット内に正しい参照スペクトルがあり、88%の確率でトップヒットにもなりました。予測データベースを実測の重複レコードで検索した場合、97%の確率で上位 10 ヒット内に正しく対応する予測スペク

トルがあり、73%の確率でトップヒットにもなりました。この結果は、実測データが依然として予測データよりも優れている一方、予測データは未知の物質を分類し、定義された化学空間内で表現される化合物の数を増やすための適切なツールとなり得ることを示しています。

2 回目の検証テスト

2 回目の検証テストは、同じパラメータを使用して実施しましたが、1 回目のテストが正確であることを確認するために、わずかに大きなデータセットを使用しました。モデルテストセットを 5%から 10%に拡張し、より多くの重複レコードを分析に使用できるようにしました。

実測のテストセットデータベースで検索された実測の重複レコードにより、100%の確率で上位 10 ヒット内に正しい参照スペクトルがあり、98%の確率でトップヒットになりました。予測データセットで検索された実測の重複レコードは、85%の確率で上位 10 ヒット内にあり、58%の確率でトップヒットになりました。2 回目の検証テストは、実測データが元の結果と比較して非常に優れたパフォーマンスを示しており、元の検証テストの結果を反映しています。予測データは実測データを下回りましたが、依然として高いレベルでした。

2 回目のテストの結果を調査したところ、75 を超える良好な HQI 値を示したものが多くありましたが、構造タイプが原因で上位 10 ヒットには含まれませんでした。例えば、予測された長鎖炭化水素は、赤外分光法における多くの同様の構造の炭化水素スペクトルによって、ヒットリストにおいて圧倒された可能性があります。長鎖炭化水素のスペクトルの類似性は十分に高いため、重複レコードのわずかな違いにより、予想されるヒットリストの順序が非常に簡単に変わってしまう可能性があります。1 回目のテストの重複レコードが非常に多様で、2 回目のテストの重複レコードが酷似している場合、ヒットリストの位置が低下するもう一つの可能性は、重複レコードの選択である場合があります。

最適化補正がどちらのデータセットの結果にも強く影響しているかどうかを確認するために、最適化補正を使用せずに同じデータセットでも検証テストが実施されました。このテストでは、一部の結果が影響を受けましたが、全体的なヒットリストの順位スコアは、最適化された補正を有効にしたテストと比較的同じであることが示されました。実測データに基づいて検索された重複レコードは、上位 10 ヒットの正しい結果に関しては 91%に低下し、トップヒットでは同様に 88%のスコアでした。レコードは、94%の確率で上位 10 ヒット内であり、64%の確率でトップヒットになりました。ここで重要なことは、最適化補正を無効にすると、すべての目的の実測データは同じようにテストされたのに対し、予測データは補正を有効にするとより良好な結果が得られたということです。ただし、これはトップヒット結果にのみ影響し、上位 10 のヒット結果は比較的同じままでした。

結果表:

	平均ヒットリストの位置	上位 10 ヒット率
検証テスト 1: 実測のテストセットで検索された重複レコード	3.97	94%
検証テスト 1: 予測テストセットで検索された重複レコード	2.94	97%
検証テスト 2: 実測のテストセットで検索された重複レコード	1.11	100%
検証テスト 2: 予測テストセットで検索された重複レコード	9.45	85%

結論

Wiley Database of Predicted IR Spectra のライブラリが現場でどのように機能するか実際にテストする必要がありました。本検証試験は、最終製品の外部テスト、つまり、テストセットの計算による評価を超えたテストとなるように設計されました。理想的なテストは、構造が完全に一致するスペクトルライブラリに対してスペクトルを検索することで生成されたヒットリストの分析でした。次に、テストセットデータを使用して外部自動検証テストを作成しました。このテストでは予測結果が実測の結果とほぼ同様に機能することが示されました。予測データベースは実測データと同等の検索を行い、1 回目のテストではいずれも約 95%の確率で上位 10 件のヒットリスト結果に正しく一致し、2 回目のテストでは両方とも 85%超でした。実測データが依然として優れていることは明らかなものの、予測データは実測データと同様のレベルで機能していることが分かります。

これらの結果は、*Wiley Database of Predicted IR Spectra* 内の予測データセットが、利用可能な化学物質を拡大するための実測の参照データセットの貴重な補完物として有用であることを証明しています。予測ライブラリは、特に Wiley の広範なスペクトルコレクション（プレミアム Sadtler スペ

クトルを含む）といった、広範かつ高品質な実測の参照データセットから構築された場合は、実測のデータセットに迫るパフォーマンスレベルを示すため、分析ワークフローに効果的に統合できます。

参考文献

1. Wiley’s Database of Predicted IR Spectra. John Wiley and Sons, Inc. 2024 https://sciencesolutions.wiley.com/machine-learning-and-prediction-at-wiley-science-solutions/
2. Ho, S.Y., Phua, K., Wong, L., & Goh, W.B.G. Extensions of the external validation for checking learned model interpretability and generalizability. Patterns. 2020, Volume 1 Issue 8. DOI: https://doi.org/10.1016/j.patter.2020.100129
3. KnowItAll Analytical Edition 2024, Release 24.0.59.0. https://sciencesolutions.wiley.com/knowitall-analytical-edition-software/, John Wiley and Sons, Inc. 2024
4. KnowItAll IR Spectral Database Collection. John Wiley and Sons, Inc. 2024 https://sciencesolutions.wiley.com/solutions/technique/ir/knowitall-ir-collection/