ACM RecSys Challenge 2022 のインテル® SIHG4SR ソリューション

インテルの推薦システムが上位入賞

Chendi Xue インテル コーポレーション マシンラーニング・エンジニア
Xinyao Wang インテル コーポレーション AI フレームワーク・エンジニア
Zhou Yu インテル コーポレーション AI フレームワーク・エンジニア
Jian Zhang インテル コーポレーション AI ソフトウェア・エンジニアリング・マネージャー
Ke Ding インテル コーポレーション主任エンジニア AI 応用研究
Eric Anderson インテル コーポレーション AI フレームワーク・エンジニア
Rita Brugarolas Brufau インテル コーポレーション ソフトウェア・ソリューション・エンジニア

課題

米計算機学会（ACM: Association for Computing Machinery）の推薦システム・カンファレンスでは、毎年 RecSys Challenge と呼ばれるデータサイエンス・コンテストを開催しています。インテルは、このコンテストに 2 年前から参加しています。2022 年、インテルは SIHG4SR: Side Information Heterogeneous Graph for Session Recommender（英語）ソリューション（『ACM Digital Library』に掲載）を発表し、最終リーダーボードで 4 位に入賞しました（チーム名「MooreWins」）。

© Intel Corporation. Intel、インテル、Intel ロゴ、その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。
コンパイラーの最適化に関する詳細は、最適化に関する注意事項を参照してください。

RecSyc Challenge 2022（英語）は Dressipi（英語）によって企画され、課題はファッション推薦でした。提供されたデータセット（英語）には、18 か月間にわたる 110 万件の匿名化されたオンライン小売セッションが含まれていました（**図 1** と**図 2**）。各セッションには、購入されたアイテムをラベルとする、1 日の顧客の行動が含まれており、記述的なファッション特性のセットに基づいて、顧客がどのアイテムを購入するかを予測することが課題でした。コンテストに参加したソリューションは、MRR（Mean Reciprocal Rank）スコアによって順位が付けられました。具体的には、各セッションにおいて、購入される可能性のあるアイテムのリストに対して、実際に購入されたアイテムがより上位に表示されれば、MRR は高くなります。逆に、アイテムが予測リストの下位に表示された場合、MRR は低くなるか、ゼロになります。例えば、予測リストが [item_2, item_3, item_1, item_0] で、item_3 が実際に購入されたアイテムの場合、$\text{MRR}=\frac{1}{\text{item}_3\text{ のランク}} = \frac{1}{2}$ になります。

図 1. オンライン・ショッピング・セッションの例

図 2. アイテム特徴の例

© Intel Corporation. Intel、インテル、Intel ロゴ、その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。
コンパイラーの最適化に関する詳細は、最適化に関する注意事項を参照してください。

インテルのソリューション

SIHG4SR ソリューションは、グラフ・ニューラル・ネットワーク（GNN）を使用して各セッションから学習し、閲覧したアイテムと最終的に購入したアイテムの間のつながりを明らかにします(**図 3**)。まず、各セッションを DGL(Deep Graph Library)（英語）を使用して実装したグラフで表現し、それをネットワークに投入します。このモデルは 3 つの段階に分けることができます。第 1 段階では、入力グラフをアイテム、特徴、カテゴリーの 3 つの学習可能な埋め込みテーブルにエンコードします。第 2 段階では、特徴とカテゴリーをアイテムの埋め込みと同じ形状に再整形し、アテンション読み出し層を適用して特徴をデコードします。そして、第 3 段階では、アイテムとサイド情報を基に 2 つの独立した予測を 1 つに融合して、アイテムの最終予測を行います。

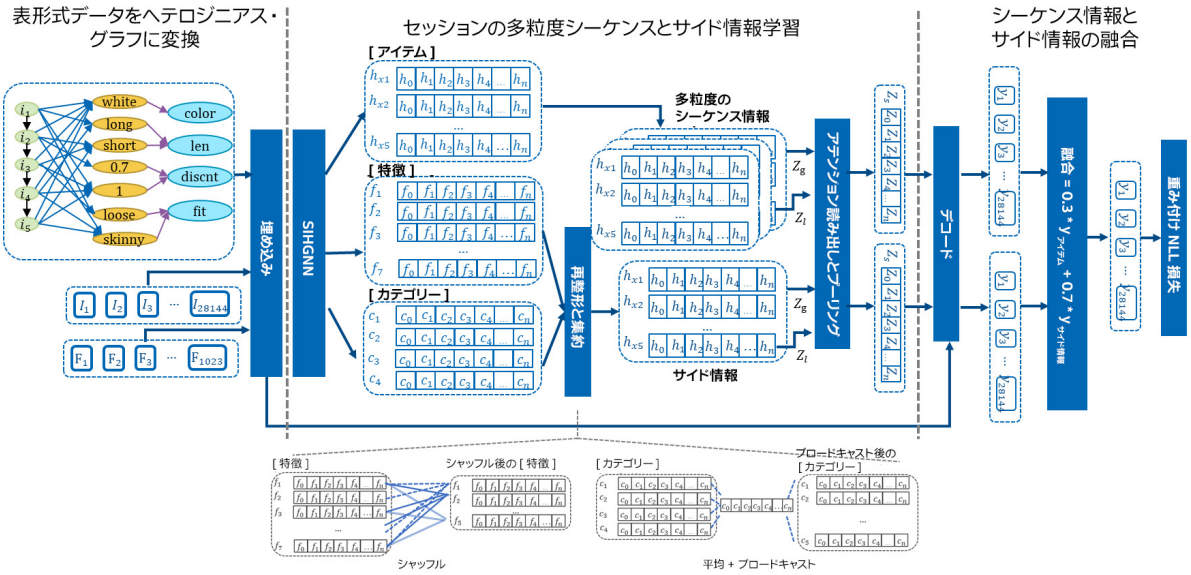

図 3. SIHG4SR アーキテクチャー

この新しいモデルは、ヘテロジニアス・グラフを利用して各セッションを表現し、新しい集約手法でサイド情報とシーケンス特徴を融合します。サイド情報としてアイテムの特徴を組み込みユーザーの意図を有効活用するこのソリューションは、再帰的特徴量エンジニアリング、2 段階トレーニング手法、マルチレベルの組み合わせ、ハイパーパラメーター・チューニングにより、MRR 0.20762 を達成し、総合 4 位を獲得しました。

このソリューションでは、表形式のセッションデータをグラフ化します（**図 4**)。各セッションは 1 つのグラフに変換され、合計 110 万個のグラフが作成されます。各グラフには 3 種類のエッジが存在します。閲覧順にアイテムをつなぐエッジ、アイテムとその特徴をつなぐエッジ、そして、特徴と対応するカテゴリーをつなぐエッジです。アイテムの順序は推薦において重要です。閲覧履歴を学習することで、グラフ・ネットワークは異なるパターンを検出し、最終的に次にクリックされるアイテムや購入されるアイテムを予測できます。さらに、各アイテムの記述的特徴と対応するカテゴリーにより、この豊富なコンテキストを利用してユーザーの意図を有効活用できます。ここでは、これらのアイテムの特徴やアイテムの特徴カテゴリーをサイド情報と呼びます。

© Intel Corporation. Intel、インテル、Intel ロゴ、その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。
コンパイラーの最適化に関する詳細は、最適化に関する注意事項を参照してください。

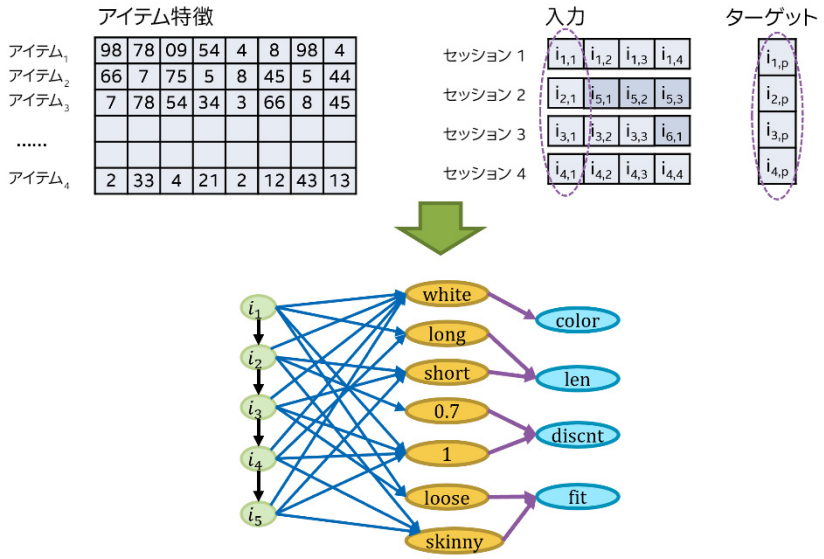

図 4. 表形式データからヘテロジニアス・グラフへの変換

モデル

このソリューションは、アイテム情報とサイド情報を集約し、異なる予測を融合し、重み付き損失関数で精度を向上する新しい方法を提案します。

アイテム情報とサイド情報の集約：特徴とカテゴリーの埋め込みの形状を統一するため、2 つの方法を採用しています。特徴の埋め込みについては、同じアイテムにつながる特徴の埋め込みを、新しい埋め込みベクトルとして合計する「シャッフル」手法を使用します。カテゴリーの埋め込みについては、まず特徴カテゴリーの埋め込みの平均を計算し、次にブロードキャストしてアイテムの埋め込みに再整形する「平均 + ブロードキャスト」手法を使用します。

アイテム予測とサイド情報予測の融合：ユーザーの嗜好を表す局所表現 z_1 と大域表現 z_g を生成し、この 2 つの表現を固定比率 $z_s = \alpha z_1 + \beta z_g$. で結合します。

重み付き損失：重み付き負対数尤度（NLL）損失関数を使用して、古いセッションよりも新しいセッションに注目します。

© Intel Corporation. Intel、インテル、Intel ロゴ、その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。
コンパイラーの最適化に関する詳細は、最適化に関する注意事項を参照してください。

特徴量エンジニアリングによる精度向上

すべての特徴がモデルに貢献するわけではないので、グラフの構築に有用な特徴を選択することは非常に重要です。提供されたアイテム特徴には、74 のカテゴリーと 904 の一意の「カテゴリー:値」のペアが含まれています。例えば、特徴カテゴリー 3 が「色」の場合、カテゴリー 3 に属する特徴の値は、「赤」や「白」、あるいは別の色である可能性があります。もし、あるカテゴリーが履歴で一貫性のない情報を提供したり、ほかのカテゴリーと矛盾するメッセージを提供する場合は、グラフから除外する必要があります。さらに、提供された情報を基に、新たな有用な特徴を生成することもできます。このソリューションでは、個々の特徴カテゴリーを除外するテストを 74 回行ってから、過去に検出された有害な特徴に応じて、特徴カテゴリーの組み合わせを除外するテストを 8 回行いました。さらに、閲覧と購入のタイムスタンプ、過去のアイテム閲覧履歴、購入回数を解析することで、予測に有益な 3 つの新しい特徴を作成しました。

インテル® End-to-End AI Optimization Kit

インテル® End-to-End AI Optimization Kit（英語）は、一般的なハードウェア上でハイパフォーマンスかつ軽量なモデルを実現する、エンドツーエンドの AI 最適化を行う構成可能なツールキットです。データ処理、トレーニング、モデル最適化、推論など、AI パイプラインの各段階を最適化することを目的に開発されました。いくつかのコンポーネントが含まれています。RecDP は大規模データ処理を行うシンプルな API を提供し、SDA（Smart model Democratization Advisor）は SigOpt（英語）でモデルを最適化してチューニングする組み込みの知識を提供し、トレーニングなしのアーキテクチャー検索キット（DE-NAS）はコンパクトなニューラル・ネットワークを構築し、Model Adapter はトレーニング前の知識を簡単に新しいモデルへ転移させます。

SDA によるハイパーパラメーター・チューニング

最適なモデル・アーキテクチャーを得るため、ハイパーパラメーター・チューニングを行い最適なパラメーターを選択しました。手動のハイパーパラメーター・チューニング（HPO）には専門知識が必要であり、通常、大規模かつ再帰的な試行の間にエラーが発生します。そこで、SIHG4SR では SDA を利用してハイパーパラメーター・チューニングを行いました（**図 5**）。

図 5. インテル® End-to-End AI Optimization Kit によるハイパーパラメーター・チューニング

Model Adapter による 2 段階トレーニング

各セッションには 1 つの購入アイテムと複数の閲覧アイテムが含まれるため、Model Adapter を使用して転移学習ベースの 2 段階トレーニングを行い、データを深く掘り下げます（**図 6**）。第 1 段階では、ネクストクリックの問題に注目します。このネットワークは次に閲覧されるアイテムを正しく予測しようとします。これにより、データサイズを 5 倍に増強し、モデルを大幅に強化できます。第 2 段階では、ネクストクリック問題から、小さなデータセットで十分な予測精度を持つモデルを生成できる購入アイテム予測問題へと知識を転移させます。

© Intel Corporation. Intel、インテル、Intel ロゴ、その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。
コンパイラーの最適化に関する詳細は、最適化に関する注意事項を参照してください。

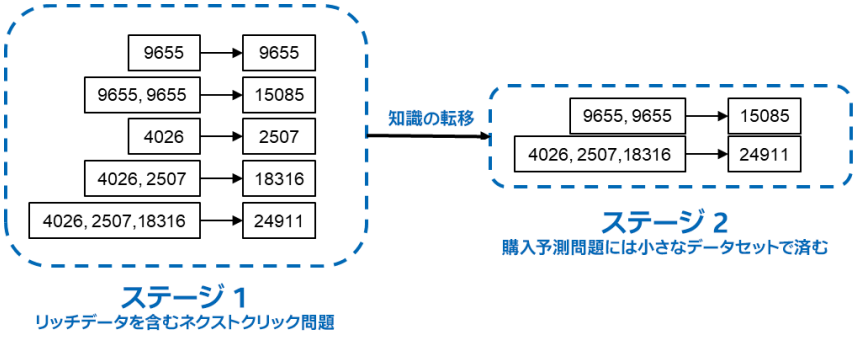

図 6. Model Adapter による転移学習ベースの 2 段階トレーニング

複数のモデルを組み合わせる

予測能力を向上させるため、SIHG4SR アーキテクチャーをベースに複数の派生モデルを構築します（**図 7**）。トレーニング手法、データ・サンプリング、ノード設定、エッジドロップ手法、読み出しと融合の設定を調整して、サブモデルを作成します。最終的に、SIHG4SR_V1、SIHG4SR_V2、SIHG4SR_V3、SIHG4SR_V3_enhanced と名付けた 4 つのサブモデルを作成しました。2 レベルの組み合わせを使用して、4 つのモデルからの予測値を加重和することで、単一モデルの一般化誤差を大幅に低減できます。

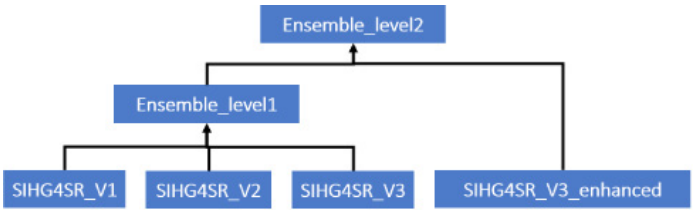

図 7. モデルの組み合わせ

まとめ

RecSys Challenge 2022 はセッションベースのレコメンダー問題で、参加者には匿名化されたデータセットが提供されました。セッションデータは表でもグラフでも表現可能であり、ディープラーニング・モデルのほうがマシンラーニング手法よりも優れた予測能力を示しました。インテルのソリューションは、4 つの重要な要因により、MRR 0.20762 を達成し、総合 4 位を獲得しました。

- SIHG4SR : セッションベース・レコメンダー問題のサイド情報を融合するヘテロジニアス GNN ベースのモデル
- グラフ内のノード / エッジを追加 / 削除して精度を向上する特徴量エンジニアリング手法
- ハイパーパラメーター・チューニングと転移学習を行うインテル® End to End AI Optimization Kit（英語）
- 複数のモデルを活用するマルチレベルの組み合わせ

© Intel Corporation. Intel、インテル、Intel ロゴ、その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。
コンパイラーの最適化に関する詳細は、最適化に関する注意事項を参照してください。