D-012

推薦システムにおける推薦理由の説明可能性に関するサーベイ

A Survey of Explainable Recommender System

松島 ひろむ [1]　森澤竣 [2]　石山琢己 [2]　山名 早人 [3]
Hiromu Matsushima　Shun Morisawa　Takumi Ishiyama　Hayato Yamana

1. はじめに

推薦システムは Web ページ、動画配信サイト、音楽アプリなど様々な場面で利用されている。しかし、推薦システムの内部でどのように推薦アイテムを選んでいるのかはブラックボックスとなっており、推薦システムの透明性、 説得力、 有効性、 信頼性、 満足度を向上させるために利用者に対して推薦理由を提示することが求められている。

推薦理由を提示する研究は、モデル内在型とモデル独立側に分類できる。モデル内在型は、元来の推薦モデルに改変を加え推薦理由を抽出しようとするものである。一方、モデル独立型は元来の推薦モデルをブラックボックスとして考え、元来の推薦モデルへの入力と出力のペアを外部から観測（学習）することで、元来の推薦モデルとは独立に説明を生成しようとする手法である。近年、深層学習の発展により推薦モデルの内部はよりブラックボックス化されており、こうした深層学習モデルに対しては、モデル独立型の手法がとられる場合が多い。

推薦の説明文を提供できるようになると、 ユーザやシステム設計者は、特定のアイテムが推奨される理由を知ることができる。さらに、 システム設計者がシステムのデバッグを行う際にも役立つ [1] 。近年多くの説明可能な推薦手法が提案され、 実際のシステムに適用されている。

本稿では、説明可能な推薦システムの研究動向と、現状の研究課題についてまとめる。本稿は、以下の構成をとる。2 節で説明可能な推薦について概要を述べた上で、3 節で調査方法 についてまとめ、4 節で調査結果を報告する。5 節で 調査結果を踏まえた実用上の課題に対する既存研究について紹介し、 最後に 6 節でまとめる。

2. 説明可能な推薦

本節では、 説明可能な推薦の概要と課題について述べる。

2.1 説明可能な推薦の概要

説明可能な推薦とは、アイテムの推薦と同時に推薦理由を提示する推薦アルゴリズムを指す。 説明可能な推薦は、推薦アイテムを提供するだけでなく、 アイテムが推薦される理由を明確にする。これにより、推薦を利用するユーザまたはシステム設計者に対する推薦システムの透明性が向上する。また、推薦を利用するユーザに対する説得力、 有効性、 信頼性、 ユーザ満足度も向上する。 例えば EC サイトでは、ユーザの好みに合った説明文はより多くの購買に繋がることが報告されている [9] 。システム設計者は推薦アルゴリズムを診断、 デバッグ、 および改良するのが容易になる [1] 。

推薦システムにおける推薦理由の生成手法には、モデル内在型とモデル独立型の２つのアプローチがある。モデル内在型では、 推薦モデル内部の学習パラメータが解釈可能なアルゴリズムを 採用することによって，説明を可能とする。一方、モデル独立型は、推薦アイテムの決定メカニズムがブラックボックスであり、元来の推薦モデルに手を加えずに推薦理由を提示したい場合に使用する。元来の推薦モデルとは別のモデルで入力データと出力データ（推薦アイテム）を観測することで、説明文を生成する。[2]

2.2 実用上の課題

2.2.1 推薦の精度と解釈性のトレードオフ

モデル内在型の解釈性と推薦の精度には概してトレードオフの関係が存在する [1] 。説明可能な推薦システムでは、解釈性を上げて具体的な説明を生成するためアイテムの決定メカニズムの精度が犠牲になる [1] 。 近年開発が進んでいる深層学習は高い精度で推薦をすることができる一方、内部の隠れ層でどのような計算が行われているかはブラックボックスであり、解釈が難しい。また、解釈を可能にするためにモデルに手を加えると精度を下げてしまうことがある。

一方、モデル独立型の場合には、元来の推薦モデルへの入力とその出力のペアを解釈性の高い学習器で学習する。このため、元来の推薦モデルの精度を保ったままで推薦システムを構築できる。一方で、入力と出力の関係を改めて学習することになるため、元来の推薦モデルの説明理由を十分に説明することが困難である。

2.2.2 適切な説明文の生成

説明文は 1 行ほどの説明から 100 字以上の具体的な説明まで幅広いパターンが存在する。既存の推薦システムは「あなたに似たユーザが気に入っています」と言った簡潔な推薦文は生成できるが、推薦システムの透明性を高めるためには、より具体的な説明文を生成する必要がある [4] 。例えば、説明文の信頼性を上げるために、ユーザのアイテムへの好みに沿ったアイテムの特徴やユーザの現在位置や時間、ユーザに似た人のアイテムへの好みなどを盛り込むことも重要である。

[1] 早稲田大学 基幹理工学部 情報理工学科
Undergraduate School of Fundamental Science and Engineering, Waseda University
[2] 早稲田大学大学院 基幹理工学研究科情報理工・情報通信専攻
Graduate School of Fundamental Science and Engineering, Waseda University
[3] 早稲田大学理工学術院
Faculty of Science and Engineering, Waseda University

医療や薬の推薦では、推薦アイテムや説明文の間違いがユーザにとっての危険を招く恐れがあるため、正しく信頼性のあるアイテムや説明を生成しなければならない [5]。

なお、ユーザの評価データや閲覧履歴が少ない場合、ユーザのプロフィール情報などの限られたデータを元に推薦を行わなければならないコールドスタート問題が起こる。データが少ない場合、説明文も限られた情報から生成しなければならないため説明文の品質が下がる。

2.2.3 評価指標や実装に関して

説明可能性に関しての評価指標は、明確に定まっておらず、既存手法との比較による評価、あるいは絶対値による評価を被験者アンケートにより実施するのが一般的である。

3. 調査方法

本稿では、 以下の手順に分割して研究調査を行った。

1. 調査目的を設定する。
2. 論文データベースを用いてキーワード検索を行う。
3. 検索結果から、 調査目的に合致する論文を手作業で抽出する。

3.1 手順 1: 調査目的の設定

本稿は、 説明可能な推薦の研究動向と現状の研究課題について調査することを目的としている。 そこで、 本研究の対象とする論文を、 以下のように定めた。

・国際学会あるいは論文誌で発表されている。
・推薦システムの説明可能性に関する論文である。

ここで推薦システムの説明可能性とは推薦システムにおいて高品質の推薦だけでなく推薦した理由の説明も生成できるかどうかを指す。

3.2 手順 2: 論文データベースを用いたキーワード検索

論文データベースを用いて、 キーワード検索を行った。調査対象の論文を網羅的に収集するため、 論文データベースとして ACM DIGITAL LIBRARY、 IEEE Xplore を選定した。検索条件としては、 以下の項目を指定した。 ただし、キーワード 検索の対象は、論文の抄録である。検索を実施した日付は、 2021 年 3 月 25 日である。

・キーワード: “explainable” AND “recommend”
・出版年: 2016 年~2021 年
・文献種別: Conference Paper または Article

3.3 手順 3: 調査対象論文の抽出作業

手順 2 で得られた検索結果から、本稿の対象と する論文を選定した。選定は、手順 1 で定めた条件をもとに、タイトルと抄録と結論の内容を確認する形で行った。

4. 調査結果概要

手順 1 及び手順 2 の結果、 ACM DIGITAL LIBRARY では 56 件、IEEE Explore では 41 件の論文がヒットした。手順 3 を経て、ACM DIGITAL LIBRARY では 38 件、IEEE Xplore では 34 件に絞り込んだ 。図 1 に各年で出版された説明可能な推薦システムの論文数の推移を示す。

図中、モデル内在型は、CF(Collaborative Filtering)、DNN(Deep Neural Network)、 Knowledge Graph(KG)、Topic、Others(モデル内在型)に分類した。モデル独立型は LIME(Local Interpretable Model-agnostic Explanations) と Others(モデル独立型) に分類した。

CF は主にユーザのアイテム評価データを用い、類似するユーザの好みを元に推薦をするモデルである。DNN は深層学習を用いた推薦モデルであり、画像解析等が含まれる。Knowledge Graph はアイテムとアイテムの周辺情報をノードと、ノード間の関係性を表すエッジで表したグラフを用いたモデルである。Topic モデルはレビュー文の解析を行い重要な単語を見つけることやレビュー文に含まれているポジティブやネガティブな単語を元に感情分析を行うモデルである。LIME は特徴空間内の 局所的な部分に対してシンプルな回帰モデルによる学習を行うことによって，回帰係数を特徴量の重要度として解釈する手法である。

図 1 よりモデル内在型を用いた手法（DNN や Knowledge Graph、Topic モデルなど）が、モデル独立型よりも増加傾向にあることが分かる。なお、DNN や CF は、Knowledge Graph や Topic モデルでも用いられているため明確に推薦モデルを分けることはできなかった。今回は、複数の手法を組み合わせている場合には主に用いられている方のモデルを統計に含めた。

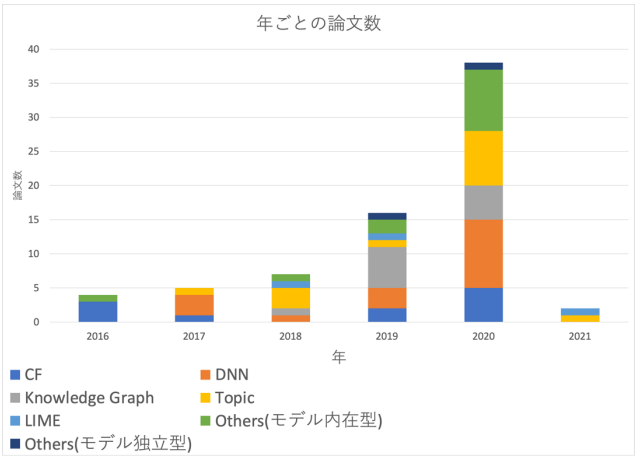

図 1 年別の論文数の推移

5. 実用上の課題に対する研究

本節では、 2 節で述べた説明可能な推薦システムの実用上の課題点について、 既存研究をまとめる。

5.1 説明可能な推薦の研究動向

5.1.1 モデル内在型

5.1.1.1 CF モデル

協調フィルタリング(CF)ベースの推薦システムは広く普及しており、「過去に同様のアイテムを使用したユーザが今後も似たようなアイテムを使用し続ける」という想定に基づいたモデルである[6] 。CF モデルは、各ユーザがいくつかのアイテムをどれだけ好きか嫌いかについての情報を提供する評価マトリックスに基づいて予測スコアを計算する。

CF は、メモリベース手法とモデルベース手法に分類できる。メモリベース手法では、評価行列から情報を直接取得し、近傍法などを用いて似ているアイテムやユーザを元に推薦アイテムを決定する。モデルベース手法は評価行列を元に事前に推薦モデルを作成しておく手法である。

メモリベース手法は、さらにユーザベースとアイテムベースの 2 種類に分類できる。ユーザベースでは、ターゲットユーザと同様の興味や経験を共有するユーザのグループを分析し、グループが一般的に好むアイテムを推奨する。アイテムベースでは、ユーザが過去に気に入ったアイテムのリストとの類似性が高いアイテムを推奨する。

モデルベース手法では、評価マトリックスを元に事前に行列因子分解（MF: Matrix Factorization）や回帰などを用いて推薦モデルを構築する。モデルベースのアプローチの最も一般的な実装は、行列因子分解である。行列因子分解では、アイテムに対するユーザの評価はユーザとアイテムの特徴を表す一連の潜在ベクトルを使用してモデル化される[1] 。モデルベースの方法は一般的にメモリベースより高い精度であるが、学習された潜在空間を解釈するのが容易ではないため精度をできるだけ下げずに解釈性を高めることが課題となる [4]。

Barkan ら [7] はユーザをペルソナのセットによってモデル化することで CF モデルの解釈性を高めた。それぞれのペルソナはアイテムの特徴の異なる傾向を表す。注意メカニズム [8] を使用して各ユーザのペルソナを動的に重み付けし、ユーザを複数のペルソナの重要度の組み合わせで表す。推薦アイテムはユーザとアイテムのペルソナに基づくアテンションスコアの類似性によって決定でき、最も関連性の高いペルソナに基づいて各推奨アイテムを説明する。

Sato ら [9] はアップリフトの最適化と説得力のある説明の 2 つに焦点を当てた。アップリフトは推薦によって引き起こされる購入確率の増加として定義する。[9]では、文脈を用いた新しい説明スタイルを提案している。フィールド認識因数分解マシン（FFM: Field-aware Factorization Machine） [10] を使用してユーザ、アイテム、文脈間のペアワイズ相互作用を学習する。説明は、人間が作成したテンプレートを元に「アイテム i は、コンテキスト c に適しているため、お勧めします」といったように生成する。また、7 種類の説明スタイルを用意し、説明スタイル毎の説得力と有用性を測定するために、クラウドソーシングベースのユーザ調査を実施した。

5.1.1.2 深層学習モデル

深層学習モデルは深層学習の技術を利用したモデル内在型を指す。入力特徴を元にニューラルネットワークを用いて予測をする場合どの特徴量を重視して推薦アイテムが選ばれたのかがブラックボックスとなる。ブラックボックスになることを避けるために、似ている種類の特徴量ごとに入力を分割し、それぞれの分割でニューラルネットワークによる予測を行い、最後にブースティングさせて出力することで解釈性を高めるモデルが提案されている。 [11] CNN、RNN/LSTM、注意機構(Attention Mechanism)などの深層学習を用いた技術を用いて、入力から取得した埋め込みを元に推薦アイテムと説明文を同時に出力するモデルも提案されている。畳み込みニューラルネットワーク（CNN）は、アイテムの説明文やユーザレビューなどの自然言語入力から潜在表現を学習できる。リカレントニューラルネットワーク（RNN）を用いることで、レビュー文などのアイテムに関連する文章から埋め込みを学習し、埋め込みを元に文生成を行うことができる。入力をタイムリーなシーケンスとして扱いそれぞれの表現を生成する手法である。RNN は、テキストのみに限定されるものではなく、コンテキストを考慮した推薦を行うコンテキストレコメンダーシステムでも利用される。 [14] RNN を用いたコンテキストレコメンダーシステムで推薦するアイテムは文脈に応じて選ばれる。注意機構はモデル内部の決定メカニズムを明らかにすることができる。注意機構を使用して、推薦アイテムの決定において最も重要とした特徴を検出できる。

Afchar ら [11] は解釈可能なディープニューラルネットワークを提案した。複数のニューラルネットワークを用意し、マスク制約を使用し特徴に応じてそれぞれのニューラルネットワークへデータを流し、対応するニューラルネットワークを調べることで各入力特徴の寄与を評価できるようにした。モデルへの入力空間において、入力集合の部分集合 $X = \{X_1, X_2, \dots, X_N\}$を事前定義しておき、それぞれの部分集合に入力を分割する。部分集合は例えば、ユーザのアプリケーションへのインタラクション特徴、音楽ジャンル特徴、音楽のムード特徴等である。部分集合X_i間の関連度を定義し、関連度が高いX_i同士を統合して新たな部分集合$S = \{S_1, S_2, \dots, S_H\}$を作る。部分集合$S_i$ごとにニューラルネットワークに入力を供給する。出力では、部分集合の依存関係を保持する複数のニューラルネットワークをブースティングして予測を提供する。 各入力特徴の寄与は、特徴に対応するニューラルネットワークを調べることで評価できるため解釈可能である。各部分集合に特化したサブネットワークにアクセスして、予測の起源を解釈できる。

Zarzour ら [12] は説明可能な推薦システム構築を目指して、まず説明文生成の一助とするため、畳み込みニューラルネットワーク(CNN)ベースのレビュー分類法を提案した。元々コンピュータビジョン用に設計された CNN を用いてレビュー文が肯定的か否定的かを予測できるようにした。テキストによるレビューを単語レベルで処理し疎なベクトルに変換した後、密な埋め込みデータであるレビュー埋め込み行列に変換する。レビュー埋め込みは畳み込み層、プーリング層を経て全結合層でシグモイド活性化関数を使用して重みに基づいて二項分類の決定が行われる。

Lin ら [13] はファッションアイテムの推薦で注意機構を活用することにより、画像とユーザコメントを組み合わせて利用して洋服の推薦と説明を行なった。推薦は、服のマッチングとコメントの生成の 2 段階で構成される。まず、服のマッチングでは、相互注意機構を備えた CNN を利用して画像から洋服の視覚的特徴を抽出し、トップスとボト

ムスを潜在ベクトルとして表現する。次のステップであるコメント生成では、視覚的特徴を簡潔な文に変換するためのクロスモダリティ注意機構を備えたゲート付きリカレントニューラルネットワークを提案している。クロスモダリティ注意機構を使うことで、推薦アイテムを決定するときに注意が向けられた視覚空間をテキスト空間に変換できる。

Luo ら [15] は時間軸を考慮するアスペクトレベルの説明可能な推薦を行うモデルを提案した。注意機構を用いることで、ユーザ・アイテム間の暗黙的なフィードバックから、ユーザに対する複数のアイテムのそれぞれの重要度を取得し、パーソナライズした埋め込みベクトルを生成する。LSTM ベースのネットワークを使うことでレビューから動的な埋め込みを取得できる。ユーザとアイテムそれぞれの、パーソナライズした埋め込みと動的な埋め込みを利用して協調フィルタリングによる評価予測と、ユーザの好みベクトルとアイテムの特徴ベクトルを生成する。ユーザの好みベクトルとアイテムの特徴ベクトルに含まれているアスペクト情報を利用してテンプレートを元にした説明を生成する。

5.1.1.3 ナレッジグラフモデル

ナレッジグラフは情報同士の関連性をグラフで表したものである。ナレッジグラフにはアイテムに関する関係性や周辺情報などの豊富な情報が含まれる。ナレッジグラフを扱う際は膨大なデータ量となるため計算時間を減らすための工夫が必要となる。

Wang ら [16] はナレッジグラフの埋め込み表現とアイテムの推薦に多層パーセプトロンを使用した。推薦モデルは推薦モジュールと知識グラフ埋め込みモジュールの 2 つのモジュールから成る。推薦モジュールは、ユーザとアイテムを入力として受け取り、多層パーセプトロン（MLP: Multilayer perceptron）を介してユーザとアイテムの潜在的な表現を抽出する。ユーザ・アイテムの相互作用情報の抽出では、通常ユーザ埋め込みとアイテムの埋め込みの内積を関連性の尺度として使用するが、ユーザとアイテム間の相互作用情報は必ずしも線形であるとは限らないため、多層パーセプトロンを使用して改善した。知識グラフ埋め込みモジュールは、多層パーセプトロンを介してテールノード潜在的な表現を出力する。移動距離ベースの方法である TransH は知識グラフを処理するために使用される。TransH は、ヘッドノード、ノードの関係性、テールノード間の関係を取り出す。

Huang ら [17] はユーザの動的な関心の変化を捉える逐次推薦において説明理由を提示した。テキスト、画像、構造の 3 種類のモダリティを含むマルチモーダル融合を採用することによる共同学習方法を採用した。テキスト表現は、アイテムのタイトルと説明を収集し fastText[4] を適用して抽出する。画像表現は、ImageNet[5] 事前トレーニングした AlexNet を使用して視覚的特徴を抽出する。構造表現は、外部ナレッジグラフを導入することでアイテムの構造表現を構築し、エンティティ同士の関係性をベクトル空間に構造情報を保持した状態で埋め込む。位置エンコードモジュールを備えた自己注意メカニズムを採用することで、シーケンスの長距離依存関係をキャプチャするだけでなく、効率的に学習することができる。各パスの重要度を計算しておくことで、「これまでに視聴した映画 m3 の続編であるため、映画 m5 をお勧めします。」といったパスレベルの説明可能な推薦を提供できる。

Valdiviezo-Diaz ら [18] は特徴の階層構造を組み込むことで、テンプレートと特徴量を元に説明文の生成をした。特徴階層は、アイテムの特徴を表すノードからなる木である。子は親のサブコンセプトである。例えば、木において肉ノードの子ノードは牛肉ノードである。階層構造を構築するために Microsoft Concept Graph[6] を活用する。Microsoft Concept Graph は、500 万を超える概念と 8,500 万の IsA 関係を備えた知識グラフである。IsA 関係は「猫は動物である」といった関係である。レビューの n-gram を概念グラフの概念にマッピングする。n-gram はある文章で連続する n 個の単語のまとまりを表す。概念グラフ内の明示的な特徴を再帰的に検索し、「IsA」関係を利用して階層構造を構築する。レビューで特徴が言及された回数と感情に基づいて、ユーザの各特徴への関心ベクトルとアイテムの特徴ベクトルを計算する。モデルの出力には、予測評価と関連する特徴量のサブセット E を含む。ユーザに提示する説明はテンプレートと E を元に” You might be interested in E, on which this item performs well.”と生成する。

5.1.1.4 トピックモデル

トピックモデルでは、アイテムのレビュー文に基づいて推薦の説明を行う。レビュー文を使うことで、アイテムやユーザ情報から通常では直接取得できないアスペクトを抽出し、それぞれのアスペクトの重要度を算出できる。アスペクト毎の重要度は推薦アルゴリズムで利用することができ、重要度の可視化や重要度に基づいた説明文の生成をすることで推薦の説明に役立つ。また、レビュー文に含まれている単語を元にユーザがアイテムについて肯定的か否定的かを判断する感情分析も行うことができる。

Zhao ら [19] は会話アプリケーションにおいて、説明可能な音楽の推薦を提案している。友人からおすすめされているようにユーザが感じる推薦理由を生成することで、曲のクリック率を高めることを目標とした。本目的のために、ユーザU、推薦理由Y、曲Sからなるトリプレットで構成されるデータセットを構築し、$p(Y \mid S, U)$を最大化することで特定の曲とユーザのペアにおける推薦理由を生成することを学習するシステムを構築する。おすすめの理由となるコメントを抽出し、コメントを入力として受け取り、コメントが推薦理由として使用できるかを予測する分類器をトレーニングする。次に、特定のユーザタグに関連する推薦理由を収集する。ここで、ユーザタグとは、ユーザのステータスと関心をカバーする事前定義されたキーワードのセットである。ステータスは学生などの単語、関心は民謡などの単語で表現する。ユーザタグを事前にトレーニングされた Word2Vec に投影し、コサイン類似性の観点から類似した単語を発見する。最後に注意メカニズムを用いたエン

[4] https://fasttext.cc
[5] https://image-net.org/download.php
[6] https://concept.research.microsoft.com/

コーダーデコーダーフレームワークを使用して生成確率$p(Y \mid S, U)$をモデル化する。

Zanon ら [20] は映画シナリオのレビューから抽出した、感情的なキーワード間の意味的近接性を通じてアイテムの類似性を計算するアルゴリズムを提案した。アルゴリズムはアスペクト抽出、感情分析、アスペクトランキングの 3 つの段階から成る。アイテムに関する各ユーザレビューから、関連性の高い名詞のセットを抽出する。関連性の値はカルバック・ライブラーによって取得する。カルバック・ライブラーは確率分布の差異を測る尺度である。アスペクト抽出後の第 2 段階では、Stanford Core NLP の感情分析アルゴリズムをアイテムに関するすべてのレビューに適用する。最後に、アスペクトごとの感情スコアを用いてランキング付けをする。感情スコアの順に並べられた、アスペクトの意味的近接性に基づいて映画の類似性を計算し、推薦理由を生成する。推薦する映画とユーザの映画履歴において最も関連性の高いアスペクト情報を利用してテンプレートを用いて説明を生成する。映画履歴を X、推薦した映画をYとし、それぞれのアスペクト(A , B)を利用して、「Aの単語で定義されている映画Xを見たので、同様の単語Bで定義されている映画Yをご覧ください。」と生成する。

Wu ら [21] は洋服の推薦において、テキスト情報と画像情報を共同で学習することにより正確な推薦と画像とテキストによる説明を提供した。双方向の 2 層の適応型アテンションレビューモデルを設計して、ターゲット製品に対するユーザの画像でわかる好みと画像ではわからない好みを取得する。レビュー主導の視覚的注意モデルを提案して、過去のレビューから得られたユーザの好みによって駆動されるパーソナライズされた画像表現を取得する。アテンションネットワークを介して単語と、画像の一部分を強調表示することで説明する。

Yilma ら [22] は絵画ごとの文章による描写を活用して、絵画間の潜在的な意味関係を明らかにする LDA(Latent Dirichlet Allocation)ベースのモデルを作成した。LDA では各ドキュメントを複数のトピックのセットとして特徴付ける。絵画では「宗教」、「肖像画」などのいくつかの概念の混合として説明できる。各トピックは単語の分布で表される。各ドキュメントに関連する重要な潜在トピックはドキュメントの性質を説明する。LDA モデルは、説明可能な推薦を提供しながら、絵画間の非自明な意味関係を明らかにした。ユーザの好みに一致する推薦アイテム間で共有される重要なトピックをワードクラウドによって表した。

Bai ら [23] はナレッジグラフとレビューのアスペクト情報から抽出される感情を融合して評価を予測し、パーソナライズされた説明可能な推薦をした。明示的・暗黙的なアスペクトとユーザの感情はレビューから BERT によって抽出される。評価予測では翻訳ベースの行列分解マシン（TransFM: Translation-based Factorization Machines）が用いられる。TransFM は、行列分解マシンの利点を継承し、コンテンツベースの機能を強化すると同時に逐次推薦での翻訳ベースのモデルのパフォーマンスを向上させる。推奨理由の生成ではTransformerを用いてアスペクト融合、ナレッジ融合をする。アスペクト融合は、アスペクトとアイテムタイトルを融合する。アスペクト融合の後、ナレッジ融合でナレッジグラフから得られた関連知識を融合する。最後に、双方向注意ネットワークを使用してアスペクトと極性、アイテムのタイトル、ナレッジを組み合わせる。パーソナライズされたコンテンツが豊富な推薦理由を生成する。

Suzuki ら [24] はレビューを使用してユーザの好みや商品の情報を推薦の説明文に含めた。多基準評価データとレビューを使用して説明文を生成するリカレントニューラルネットワークモデルを開発した。評価データa が与えられたときにレビューテキストr である可能性 $p(r \mid a)$を最大化するモデルを作成した。トレーニングされたモデルはテスト時に、テストデータセットの特徴量をエンコードする。長短期記憶（LSTM: Long Short Term Memory）ユニットを使用して、各タイムステップの次の単語を条件付き確率$p(r \mid a)$の最大化により予測し、文脈ベクトルを説明文にデコードする。また、入力された特徴量をより有効に活用することを目的として注意機構を利用した。

5.1.2 モデル独立型

モデル独立型は「推薦されたアイテム」と「推薦に際しモデル入力されたデータ」のペアを学習することで、推薦モデルに依存せず推薦後に説明を生成するモデルである。推薦メカニズムが複雑で推薦理由の説明の生成ができない場合にモデル独立型は有効である。単純なモデルで複雑なモデルを近似することで複雑なモデルをシステム設計者が理解するのに役立つ。説明は、推薦アイテムを決定するための正確なメカニズムに厳密に従っていない場合があるが、多くの異なる推薦モデルに適用できる柔軟性がある。

Park ら [25] は post-hoc モデル（モデル独立型のことを post-hoc モデルと Park らは呼んでいる）で生成される説明がテンプレートを元にして特定の特徴量しか利用しておらず、説明に多様性やパーソナライズ化が十分にできていない問題に対し、グラフ構造を使うことでアイテムの多様な特徴を抽出した。製品の詳細なデータは、特定の点で類似している製品を見つけるために利用する。利用可能なすべての製品情報をグラフに結合する。製品グラフは類似の製品属性を接続し、同じ属性を共有しない類似の製品を識別できる。製品グラフ、製品、推薦アイテムを考慮したスコアを設計し、説明に使用できるデータを抽出する。ユーザと推奨アイテムの両方に関してパーソナライズされた説明を生成する。

Ribeiro ら[26] は任意の学習済み推薦モデルが与えられたとき、当該推薦モデルに対する入力―出力のペアを線形回帰モデルで学習し、結果得られた各特徴の重みを重要度として推薦モデルを解釈する LIME を提案した。LIME は複雑な推論アルゴリズムに対応するために、特徴空間内の予測対象のベクトルの近傍を学習することによって、局所的に正しい説明を生成する。推論結果の説明として出力に対する各入力特徴の重要度を数値として捉えることができるので、推薦システムでは、重要度の高い特徴を説明としてユーザに提示することできる。あらゆる推薦モデルに対して、推薦の精度を落とさずに解釈をすることが可能となる。

Jiarpakdee ら [27] は post-hoc モデル 3 つと 6 つの推薦モデルを組み合わせた比較検証を行った。実験を元に LIME のハイパーパラメータを最適化した新しい post-hoc モデルを作成した。

Zhou ら [28] は、各データが独立でない場合にモデル独立型が高い忠実度で説明できない課題に対処した。同じユーザによる映画の評価や、同じ患者の ICU 滞在期間といっ

た観測値は通常相関する。LIME はあらゆるブラックボックスモデルを説明することができるが、全ての観測値が独立であることを仮定しているため、独立でない観測値を高い忠実度で説明することができない。データ同士の相関を考慮した回帰係数と個々のデータを独立と考えた場合の回帰係数の両方を統合した線形混合モデル（LMM）を開発し、データにクラスタが存在する場合でもモデル内在型の局所的な決定境界を LIME より高い忠実度で近似した。

Morisawa ら [29] は最先端の post-hoc 手法である LIME-RS において、特徴の数が増えると解釈が難しくなる問題とユーザに提供する説明の生成が考慮されていない問題に対処した。解釈モデルにおいて説明可能な特徴の最適な数を選択することにより、特徴の数が増えても推薦モデルへの忠実度を保った。LIME が解釈する重要な特徴の中から、推薦を説明するのに適さない特徴を除外して、テンプレートを用いて説明を生成する新しい手法を提案した。

5.2 入力データの種類の動向

アイテム評価データ及びレビュー文を用いた推薦で、評価において用いられているデータセットのリストを表 1 に示す。Movielens, Yelp, Amazon のデータが多く用いられている。また、レビュー文を用いたモデルでは Yelp と Amazon のデータを用いるモデルが多い。

5.3 評価手法の動向

推薦アイテムの評価手法と説明可能性の評価手法を表 2 及び表 3 に示す。説明可能性の評価は定量的な評価より実際に生成された単語や文を定性的に評価する論文が多い。

6. おわりに

本稿では、 推薦システムにおける説明可能性について調査 を行った。 本稿で示したように、推薦システムの透明性、 説得力、 有効性、 信頼性、 満足度を向上させるために利用者に対して推薦理由を提示する様々な手法が提案されている。モデル内在型の推薦理由説明では協調フィルタリング、深層学習、ナレッジグラフ、トピックモデルが用いられていた。一方、推薦モデルとは独立に推薦理由を説明するための仕組みを用意するモデル独立型では、LIME などが用いられていた。何れの手法においても、解釈性を高めることが課題となっている。今後は、 注意ネットワークなど深層学習のようなブラックボックスモデルに代わって解釈可能な推薦を提供できるモデルについて調査していきたい。

表 1 評価において用いられているデータセット

データセット	参考文献	配布元	説明
MovieLens	[4], [7], [11], [16], [17], [20], [28]	https://grouplens.org/datasets/movielens/	movielens の Web サイトから入手した映画推薦のデータ。 各ユーザは少なくとも 20 個の映画を評価している。
Yelp	[6], [18] [23]	https://www.yelp.com/dataset	ロケーションベースの SNS である Yelp ユーザによる 590 万件以上のレビュー。 ホテル、レストラン、食料品店、ガソリンスタンド等、様々な POI に関するレビュー。
LastFM	[16]	http://millionsongdataset.com/lastfm/	last.fm オンライン音楽サイトからユーザが音楽を聴いた履歴のデータ。それぞれの曲はタグと類似の曲の情報を持つ。
Amazon Toys and Games	[12], [15], [18]	https://jmcauley.ucsd.edu/data/amazon/	Amazon が提供する「おもちゃ」と「ゲーム」に関するデータ。レビュー5 件以上を持つユーザとアイテムのみを含むように事前にフィルタリングされている。
Amazon Movies	[6], [15]	https://jmcauley.ucsd.edu/data/amazon/	Amazonが提供する「映画」に関するデータ。レビュー5件以上を持つユーザとアイテムのみを含むように事前にフィルタリングされている。
Amazon Electronics	[6], [23]	https://jmcauley.ucsd.edu/data/amazon/	Amazonが提供する「電子機器」に関するデータ。レビュー5件以上を持つユーザとアイテムのみを含むように事前にフィルタリングされている。
Amazon Home	[6]	https://jmcauley.ucsd.edu/data/amazon/	Amazonが提供する「家庭用品」に関するデータ。レビュー5件以上を持つユーザとアイテムのみを含むように事前にフィルタリングされている。
TripAdvisor	[24]	http://times.cs.uiuc.edu/%7Ewang296/Data/	TripAdvisor サイトからのデータ。userID、hotelID、総合評価、多基準評価（価値、部屋、場所、清潔さ、チェックイン/フロントデスク、サービス、ビジネスサービス）およびレビュー文からなる。 評価は 0〜5。
Film Trust	[4]	https://guoguibing.github.io/librec/datasets.html#filmtrust	FilmTrustWeb サイト全体からクロールされたデータ。ユーザのアイテムに対する 35,497 件の評価と 1,853 件の信頼性評価。

BookCrossing	[4]	https://grouplens.org/datasets/book-crossing/	書籍に関する非営利活動である Book-Crossing から収集されたデータ。278,858 人のユーザによる 271,379 冊の書籍に対する 1,149,780 件の評価データ。
Yahoo Music	[7]	https://webscope.sandbox.yahoo.com/catalog.php?datatype=r	0～100 の整数スケールでのユーザのアーティストへの評価。非常に嫌われた曲には特別なスコア 255 が与えられている。

表 2 推薦システムの評価手法

評価手法	参考文献	定義式	変数の説明	説明
RMSE(Root Mean Squared Error)	[15],[18],[23]	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i' - r_i)^2}$	n:アイテム数，r_i':予測値，r_i:実測値	二乗平均平方根誤差。
MAE(Mean Absolute Error)	[4],[15]	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert r_i' - r_i\rvert$	n:アイテム数，r_i':予測値，r_i:実測値	予測値と実測値の差の絶対値を算出し、平均したもの。
HR	[6],[11],[16],[17],[21]	$HR = \frac{N_{hits}}{N}$	N:アイテム数，N_{hits}:実際に推薦できたアイテム数	ヒット率。
Precision	[4],[14],[28]	$Precision = \frac{TP}{TP + FP}$	TP(True Positive):正しいものを薦めた数 FP(False Positive): 間違ったものを薦めた数 FN(False Negative): 間違ったものを薦めなかった数	適合率。推薦リスト中のユーザが嗜好したアイテムの割合。
Recall	[4],[14],[28]	$Recall = \frac{TP}{TP + FN}$		再現率。ユーザが嗜好したアイテムのうち推薦リストでカバーできた割合。
F1	[21],[28]	$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision}$	-	PrecisionとRecallの調和平均。
AUC(Area Under Curve)	[13],[17],[27]	-	-	Recallを横軸、Precisionを縦軸に取り、得られた曲線の下部の面積。
MAP(Mean Average Precision)	[13],[17],[20]	$MAP = \frac{1}{\lvert U\rvert}\sum_{u \in U} AP(u)$ $AP(u) = \sum_{k=1}^{N} \frac{Precision@k \times y_k}{\sum_{i=1}^{k} y_i}$	U:ユーザ集合，u:ユーザ，N:アイテム数，y_k =1:上位k番目が適合アイテム，0:それ以外	AP:適合アイテムが出現した時点をそれぞれ閾値として、閾値ごとのPrecisionを算出し、Precisionの平均をとったもの。 MAP:APのユーザ平均。
nDCG	[3],[4],[6],[11],[16]	$nDCG = \frac{DCG}{DCG_{perfect}}$ $DCG = r_1 + \sum_{i=2}^{N} \frac{r_i}{\log_2 i}$	r_i:上位i番目のアイテムの評価値，N:アイテム数，$DCG_{perfect}$:正しい順序でランキング予測をできた場合の値	ユーザのアイテムへの実際の評価値を、アイムが推薦システムによるランキングの下位に行くほど大きな値で割ったものの合計値。
MRR(Mean Reciprocal Rank)	[13],[25]	$MRR = \frac{1}{\lvert U\rvert}\sum_{u \in U} \frac{1}{k_u}$	U:ユーザ集合，u:ユーザ，k_u:uへのレコメンドのうち、最初にuが嗜好するアイテムが出現した順位	ランキングで最初に適合アイテムが見つかった順位の逆数のユーザ平均。
Uplift	[9]	$Uplift = p^t - p^c$	p^t:推薦するときの購入確率，p^c:推薦しないときの購入確率	アイテムが推薦されたときとされなかったときのアイテム購入確率の差。

表 3 説明可能性の評価手法

評価手法	参考文献	定義式	変数の説明	説明
ROUGE	[13],[23]	-	n-gram:連続するn個の単語のまとまり	生成された説明文と正解（grand truth）の間で重複するn-gramの数。Recall,Precision,F1と併せて使う。
BLEU	[13],[23]	$\text{BLEU} = BP_{BLEU} \times \exp\left(\sum_{n=1}^{N} \frac{1}{N} \log p_n\right)$	BP_{BLEU}:生成文が正解文より長い場合のペナルティ，p_n:生成文中で正解文と一致したn-gramの割合	生成された説明文と正解（grand truth）の類似度を生成文が冗長な場合のペナルティと共通するn-gramの割合で算出したもの。

GOFE	[15]	$\mathrm{GOFE} = \frac{\sum_{u \in U} c^{(u)}}{C}$	$c^{(u)}$:アイテムの説明文にユーザが満足した数、C:推薦アイテム数	前アイテムに対する説明文のうちユーザが満足する説明文を生成できた割合。
Fidelity	[28]	$Fidelity = \frac{2 \times Recall \times Precision}{Recall + Precision}$, $Precision = \frac{\lvert F \cap E \rvert}{\lvert F \rvert}$, $Recall = \frac{\lvert F \cap E \rvert}{\lvert E \rvert}$	E:モデル内在型において重要とされた特徴量、F: モデル独立型において重要とされた特徴量	モデル独立型による説明がモデル内在型にどれだけ忠実であるかを、それぞれのモデルで重要視した特徴量を元に適合率と再現性の調和平均で表したもの。

参考文献

[1] Yongfeng Zhang and Xu Chen (2020), “Explainable Recommenda- tion: A Survey and New Perspectives”, Foundations and Trends® in Information Retrieval: Vol. 14, No. 1, pp 1–101, DOI: 10.1561/1500000066.

[2] S. Morisawa and H. Yamana, “アイテム推薦理由の説明のための特徴量選択手法の検証,” *Proc. DEIM2021*, pp. 1–8, 2021.

[3] J. McInerney *et al.*, “Explore, exploit, and explain: Personalizing explainable recommendations with bandits,” *In Proc. of ACM RecSys2018*, pp. 31–39, 2018, DOI: 10.1145/3240323.3240354.

[4] P. Valdiviezo-Diaz, F. Ortega, E. Cobos, and R. Lara-Cabrera, “A Collaborative Filtering Approach Based on Naïve Bayes Classifier,” *In Proc. of IEEE Access*, vol. 7, pp. 108581–108592, 2019, DOI: 10.1109/ACCESS.2019.2933048.

[5] W. Fan *et al.*, “Graph neural networks for social recommendation,” *In Proc. of World Wide Web Conf. WWW 2019*, pp. 417–426, 2019, DOI: 10.1145/3308558.3313488.

[6] O. Tal, Y. Liu, J. Huang, X. Yu and B. Aljbawi, "Neural Attention Frameworks for Explainable Recommendation," *In Proc. of* IEEE Transactions on Knowledge and Data Engineering, vol. 33, pp. 2137-2150, 2021, DOI: 10.1109/TKDE.2019.2953157.

[7] O. Barkan, Y. Fuchs, A. Caciularu, and N. Koenigstein, “Explainable Recommendations via Attentive Multi-Persona Collaborative Filtering,” *In Proc. of ACM RecSys 2020*, pp. 468–473, 2020, DOI: 10.1145/3383313.3412226.

[8] J. Chen, H. Zhang, X. He, L. Nie, W. Liu, and T. S. Chua, “Attentive collaborative filtering: Multimedia recommendation with item-And component-level attention,” *In Proc. of ACM SIGIR2017*, pp. 335–344, 2017, DOI: 10.1145/3077136.3080797.

[9] M. Sato, S. Kawai and H. Nobuhara, "Action-Triggering Recommenders: Uplift Optimization and Persuasive Explanation," *In Proc. of IEEE ICDMW 2019*, pp. 1060-1069, 2019, DOI: 10.1109/ICDMW.2019.00155.

[10] Y. Juan and C. Lin, “Field-aware Factorization Machines for CTR Prediction,” *In Proc. of ACM RecSys2016*, pp. 43-50, 2016.

[11] D. Afchar and R. Hennequin, “Making neural networks interpretable with attribution: Application to implicit signals prediction,” *In Proc. of ACM RecSys 2020* , pp. 220–229, 2020.

[12] H. Zarzour, B. Al shboul, M. Al-Ayyoub, and Y. Jararweh, “A convolutional neural network-based reviews classification method for explainable recommendations,” *In Proc. of IEEE SNAMS 2020*, pp. 1–5, 2020, DOI: 10.1109/ SNAMS 52053.2020.9336529.

[13] Y. Lin, P. Ren, Z. Chen, Z. Ren, J. Ma, and M. de Rijke, “Explainable outfit recommendation with joint outfit matching and comment generation,” *In Proc. of IEEE Transactions on Knowledge and Data Engineering 2019*, vol. 32, no. 8, pp. 1502–1516, 2018.

[14] B. Xia, Y. Li, Q. Li, and T. Li, “Attention-based recurrent neural network for location recommendation,” *In Proc. of IEEE ISKE 2017* , pp. 1–6, 2017, DOI: 10.1109/ISKE.2017.8258747.

[15] H. Luo, N. Yang, and P. S. Yu, “Hybrid Deep Embedding for Recommendations with Dynamic Aspect-Level Explanations,” *In Proc. of IEEE Big Data 2019*, pp. 870–879, 2019.

[16] Z. Wang, Y. Li, L. Fang, and P. Chen, “Joint Knowledge Graph and User Preference for Explainable Recommendation,” *In Proc. of IEEE ICCC 2019*, pp. 1338–1342, 2019, DOI: 10.1109/ICCC47050.2019.9064099.

[17] X. Huang, Q. Fang, S. Qian, J. Sang, Y. Li, and C. Xu, “Explainable interaction-driven user modeling over knowledge graph for sequential recommendation,” *In Proc. of the 27th ACM Int'l Conf. on Multimedia.*, pp. 548–556, 2019, DOI: 10.1145/3343031.3350893.

[18] J. Gao, X. Wang, Y. Wang, and X. Xie, “Explainable recommendation through attentive multi-view learning,” *In Proc. of AAAI 2019*, pp. 3622–3629, 2019, DOI: 10.1609/aaai.v33i01.33013622.

[19] G. Zhao, H. Fu, R. Song, T. Sakai, X. Xie, and X. Qian, “Why you should listen to this song: Reason generation for explainable recommendation,” *In Proc. of IEEE ICDMW 2018*, pp. 1316–1322, 2019, DOI: 10.1109/ICDMW.2018.00187.

[20] A. L. Zanon, L. Souza, D. Pressato, and M. G. Manzato, “WordRecommender: An Explainable Content-Based Algorithm based on Sentiment Analysis and Semantic Similarity,” *In Proc. of ACM WebMedia 2020* , pp. 181–184, 2020, DOI: 10.1145/3428658.3431093.

[21] Q. Wu, P. Zhao, and Z. Cui, “Visual and Textual Jointly Enhanced Interpretable Fashion Recommendation,” *In Proc. of IEEE Access*, vol. 8, pp. 68736–68746, 2020, DOI: 10.1109/ACCESS.2020.2978272.

[22] B. A. Yilma, N. Aghenda, M. Romero, Y. Naudet, and H. Panetto, “Personalised Visual Art Recommendation by Learning Latent Semantic Representations,” *In Proc. of IEEE SMA 2020*, pp. 1-6, 2020.

[23] P. Bai, Y. Xia, and Y. Xia, “Fusing Knowledge and Aspect Sentiment for Explainable Recommendation,” *In Proc. of IEEE Access*, vol. 8, pp. 137150–137160, 2020, DOI: 10.1109/ACCESS.2020.3012347.

[24] T. Suzuki, S. Oyama, and M. Kurihara, “Toward Explainable Recommendations: Generating Review Text from Multicriteria Evaluation Data,” *In Proc. of IEEE Big Data 2018*, pp. 3549–3551, 2019, DOI: 10.1109/BigData.2018.8622439.

[25] N. Park, A. Kan, C. Faloutsos and X. L. Dong, "J-Recs: Principled and Scalable Recommendation Justification," *In Proc. of IEEE* ICDM 2020, pp. 1208-1213, 2020, DOI: 10.1109/ICDM50108.2020.00151.

[26] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "" Why should i trust you?" Explaining the predictions of any classifier." *In Proc. of ACM SIGKDD 2016*, pp. 1135-1144, 2016.

[27] J. Jiarpakdee, C. Tantithamthavorn, H. K. Dam and J. Grundy, "An Empirical Study of Model-Agnostic Techniques for Defect Prediction Models," *In Proc. of IEEE Transactions on Software Engineering*, pp. 1-21, DOI: 10.1109/TSE.2020.2982385.

[28] Z. Zhou, M. Sun, and J. Chen, “A model-agnostic approach for explaining the predictions on clustered data,” *In Proc. of IEEE ICDM 2019*, pp. 1528–1533, 2019, DOI: 10.1109/ICDM.2019.00202.

[29] S. Morisawa and H. Yamana, “Faithful Post-hoc Explanation of Recommendation using Selected Features,” *In Proc. of AAAI Spring Symposia*, 2021.