

# 動画配信サービスにおける eXplainable AI を用いたレコメンドモデルの評価

明石航<sup>1</sup> 相場邦宏<sup>1</sup> 伊藤拓<sup>1</sup> 加藤剛志<sup>1</sup> 佐藤篤<sup>1</sup> 太田賢<sup>1</sup>

**概要**：EC サイトや動画配信サービスにおいて、ユーザへコンテンツを推薦する様々なレコメンドシステムが活用されている。レコメンドシステムにおける課題のひとつとして、推薦理由が不明確なことによる信頼性やユーザ満足度の低下が挙げられ、その結果 EC サイトにおける購買意欲の低下、動画配信サービスにおける視聴作品数減少に繋がる。この課題を解決する方法として、説明可能な AI (XAI : eXplainable AI) をレコメンドシステムに応用することで推薦結果と共に推薦理由をユーザに提示する方法が考えられる。本稿では、XAI を用いたレコメンドモデルのひとつである Attentive Multitask Collaborative Filtering (AMCF) において、ユーザに最適な推薦理由やバリエーションに富んだ作品を提示できているか、実際の動画配信サービス視聴ログを用いてオフライン検証を行った。

**キーワード**：レコメンド、動画配信サービス、説明可能な AI, XAI

## 1. はじめに

Amazon や Netflix, YouTube などの Web コンテンツ配信サービスでは、ユーザの興味や関心に合致するコンテンツを推薦するレコメンドシステムを活用している。レコメンドシステムを活用することで、膨大な数のコンテンツの中からユーザが自ら検索することなく興味や関心に合致するコンテンツに辿り着くことができる。商品や映像など、多種多様なコンテンツがインターネット上に溢れている現代において欠かせないシステムとなっている。レコメンドシステムにおける課題として、推薦理由が不明確なことによる信頼性やユーザ満足度の低下が挙げられる [1]。レコメンドシステムに対する信頼性やユーザ満足度の低下はサブスクリプションサービスにおける会員離脱や非アクティブ化に繋がる。

一方で、レコメンドシステムに限らず医療、金融、セキュリティなど、あらゆる場面で人工知能 (AI : Artificial Intelligence) が活用されており、関連する技術分野として説明可能な AI (XAI : eXplainable AI) が近年注目を集めている。AI には精度のみでなく、その判断に至った経緯や理由を求める場面が多く存在する。XAI はレコメンドシステムにも応用可能であり、推薦コンテンツのみでなく、なぜそれがユーザに対して推薦されたのかを提示することができる。このように推薦理由を提示することで納得感やユーザの興味を惹く情報の提示を行うことができ、前述したレコメンドシステムにおける信頼性やユーザ満足度の低下といった課題に対応することができる。

本研究では、実際の動画配信サービスの視聴ログを用い、XAI を用いたレコメンド (以下「XAI レコメンド」という。) モデルの評価を複数指標で行った。XAI レコメンドモデルのオフライン評価指標の提案数は少ないが、今回はその中でも重要と思われる指標についてオフライ

ン評価を行うことで総合的に有用性を判断した。

本稿の構成は以下の通りである。2 章では、XAI レコメンドモデルの複数の方式を概観する。次に、3 章では動画配信サービスにおいて XAI レコメンドをどのように応用するのかに関して言及する。4 章では今回のオフライン評価に用いる評価指標、対象となるデータ、モデル、評価条件、評価結果と考察を述べる。最後に 5 章にて本研究のまとめ、今後の課題を述べる。

## 2. 関連研究

XAI レコメンドモデルには複数の方式が存在し、それぞれのアルゴリズムには推薦コンテンツの精度、推薦理由の精度に特徴がある。本章では、XAI レコメンドモデルの大きな分類やそれぞれの特徴に関して述べる。

### 2.1 協調フィルタリング

協調フィルタリングは従来より実装の簡便さと精度の高さからレコメンドシステムにおいて多く採用されている手法である。また、XAI レコメンドにおいても多数モデルの提案がされている。

Zhang ら [2]は、ユーザの嗜好を推定しそれを商品の推薦文に埋め込むため、各商品のレビュー文章からユーザの関心のある商品特徴を抽出した。それにより「この製品はあなたの関心のある機能において性能が良い」、「この製品はあなたの関心のある機能において性能が悪い」といった内容をユーザに伝えることができる。結果、レコメンドシステムの信頼性向上に繋がる。Abdollahi ら [3]は、コンテンツのメタ情報を利用せずにレーティングデータのみを用い、各ユーザの類似度から、「あなたと似た多くのユーザがこの商品を購入している」といった推薦理由を提示するモデルを提案している。

<sup>1</sup> 株式会社 NTT ドコモ  
NTT DOCOMO, INC.

## 2.2 Deep Learning

現在様々な分野で飛躍的な成果を収めている Deep Learning だが、近年ではレコメンド分野での活用をされることも多くなっている。Deep Learning ベースのレコメンドアルゴリズムにおいても、近年協調フィルタリングと同じく XAI レコメンドに応用した研究が行われている。

Seo ら [4]は、Deep Learning 手法のひとつである Convolutional Neural Network (CNN) を用いて商品レビュー文章の特徴を獲得し、推薦文に応用するモデルを提案している。各ユーザの書いた商品レビュー文章、各商品のレビュー文章それぞれの特徴を学習することで、各ユーザが商品のどの特徴に着目しているかを重みによって示すことができる。Chang ら [5]は、自然言語による推薦理由の説明文を生成するアプローチを提案した。作品のトピックとなるキーワードを抽出するモデルとクラウドソーシングによる説明文の作成により、ユーザの行動から適切な説明文を生成、ユーザに提示した。

## 2.3 その他の手法

協調フィルタリング、Deep Learning 以外にも XAI レコメンドモデルは多く提案されている。

He ら [6]はユーザと商品の2つの関係のみでなく、商品レビュー文章から抽出したキーワードも加えて3つの関係をグラフ構造化することで各ユーザへ推薦商品と重要なキーワードを決定するモデルを提案した。Davidson ら [7]は、YouTube におけるユーザの視聴履歴から視聴された動画ペアをカウントし、関連性のスコアを決定するためにユーザの評価などの行動情報を用いている。動画を推薦する際には関連性が高いと判断されたユーザの行動を共に提示する。

## 2.4 関連研究まとめ

以上で述べた通り、XAI レコメンドには様々なアプローチが存在する。一般的に XAI モデルにおける推論精度と解釈性の間にはトレードオフの関係があると言われており、Deep Learning など複雑な手法を用いて精度を向上させるとモデル解釈性は低下する [8]。2.2 節で述べた通り、レコメンドアルゴリズムにおいても複雑性の高い手法が提案されているが、単純なアルゴリズムでもパラメータを最適化することで高い精度を担保できることがわかっている [9]。その中で協調フィルタリングは従来より継続的に用いられているレコメンドアルゴリズムであり、精度が高く複雑性が低いためモデルの解釈が比較的容易である。また、関連研究では多くのアプローチで各ユーザの商品レビュー文章、各商品のレビュー文章からキーワードを抽出している。しかし、レビューを投稿しないユーザも多く、そのようなユーザには適切な推薦、推薦理由の提示ができない。また、レビュー投稿機能自体備わっていないサービスもあり、レビュー文章を用いることが前提のモデルは汎用性に欠ける。

そこで、本研究では協調フィルタリングベースの XAI レコメンドモデルである Attentive Multitask Collaborative Filtering (AMCF) [10]を評価対象モデルとして用いる。このモデルはユーザの履歴とジャンルなどのコンテンツメタ情報を入力として、各ユーザがどのメタ情報を好むかを推定する。さらに、協調フィルタリングベースのモデルの中でも推薦内容の精度が高く、推薦理由の精度も高い。また、入力データに商品レビュー文章は必要なく、一般的に付与されているメタ情報でモデル構築が可能であるため汎用的である。

## 3. 動画配信サービスでの XAI レコメンド応用

1 章で述べた通り、動画配信サービスにおいてもレコメンドシステムは多く活用されている。図 1 のように各ユーザの興味を惹くコンテンツをリスト形式で表示し、ユーザはその中から興味を持ったコンテンツを選び視聴する。また、幅広い作品を推薦することでユーザが出会ったことのないコンテンツにも出会うことができ、セレンディピティ向上に繋がる。



図 1 動画配信サービスにおける通常のレコメンドイメージ

一方で、1 章で述べたように、レコメンドシステムには推薦理由が不明確なことによる信頼性やユーザ満足度の低下といった課題がある。その課題は動画配信サービスでも同様のことが言え、なぜ推薦されたかが分からないコンテンツは視聴に繋がりにくい。実際に動画配信サービスに対してユーザから「なぜか〇〇というドラマを勧められる」、「なぜおすすめにこの作品が出てくるのか」といった、推薦コンテンツに理由を求める声が上がっている。

動画コンテンツには「SF」や「アクション」などの作品ジャンル情報、作品の出演者情報など、各作品それぞれにユーザの興味を惹くようなメタ情報が付与されている。図 2 のように、コンテンツの特徴を表すメタ情報を推薦文に加えることで推薦コンテンツに対する納得感向上やユーザの興味を惹くことが可能となる。例えば、「あなたにおすすめの SF 作品」や、「あなたにおすすめの俳優 A が出演する作品」など、ユーザに推薦した理由としてコンテンツメタ情報を提示する利用方法が考えられる。このように、XAI レコメンドを活用することで、ユーザの興味を惹くコンテンツのみでなく、メタ情報まで機械的に決定することが可能となる。

以上より、XAI を動画サービスのレコメンドシステムに

適用する際の要件として下記が挙げられる。

- ① レコメンド性能
  - 推薦コンテンツがユーザの興味に合致している
  - コンテンツが網羅的に推薦されている
- ② 推薦理由の妥当性
  - 推薦理由がユーザの興味に合致している
  - 推薦理由と推薦コンテンツに矛盾がない



図 2 動画配信サービスにおける XAI レコメンドイメージ

## 4. 評価

本章では, XAI レコメンドモデルの評価方法, 評価条件, 評価結果と考察に関して述べる。

### 4.1 評価方法

本研究では, XAI レコメンドモデルを総合的に評価するため, 複数指標を用いてオフライン評価を行った。3 章で述べた通り, 図 3 のように XAI レコメンドの評価観点として大きく分けて 2 つ挙げられる。1 つがレコメンドモデルとしての性能, もう 1 つが推薦理由の妥当性である。

レコメンドモデルとしての性能には, 推薦したコンテンツに各ユーザが興味を持つかという推薦コンテンツの精度, 各ユーザにバリエーション豊富なコンテンツを提示できているかというカタログカバレッジが評価指標として考えられる。各ユーザの興味を惹くコンテンツを推薦しつつ, 膨大な数のコンテンツからユーザが新たな発見を行えるよう, コンテンツを幅広く推薦することがレコメンドシステムとして重要である [11]。

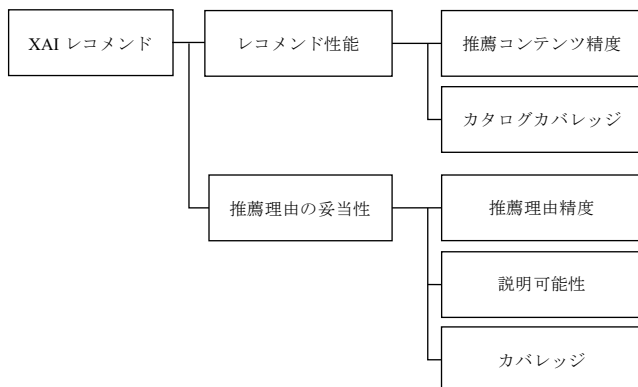


図 3 XAI レコメンドモデルに求められる要件

推薦理由の妥当性には, 各ユーザがその推薦理由に興味を持つかという推薦理由の精度, 推薦理由と推薦コンテンツの合致度によって表される説明可能性, 幅広い推薦理由を提示できているかというカバレッジが評価指標として考えられる。各ユーザの興味を惹き, 説明可能かつ幅広く提示できることが推薦理由を提示する上で求められる。

### 4.1.1 レコメンドモデル評価

#### (1) 推薦コンテンツ精度

推薦コンテンツ精度は, 学習期間のデータで学習したレコメンドモデルを用いて推薦したコンテンツを評価期間中にコンバージョンしているユーザの割合で定義する。

$$Content Accuracy@N_{recommend} = \frac{|U_{cv}|}{|U|} \quad (1)$$

$N_{recommend}$  は各ユーザに対する推薦コンテンツ数,  $U$  はユーザの全体集合,  $U_{cv}$  は  $N$  個推薦したコンテンツのうち 1 個でもコンバージョンしたユーザの集合を表す。この指標を用いることで, 推薦コンテンツが各ユーザの興味を惹くかどうかを評価期間中のコンバージョンから推測することができる。

#### (2) カタログカバレッジ

カタログカバレッジは, 全コンテンツ中の各ユーザに一度でも推薦されたコンテンツの割合で定義する。

$$Catalogue Coverage@N_{recommend} = \frac{|C_{recommend}|}{|C|} \quad (2)$$

$C_{recommend}$  は各ユーザに 1 度でも推薦されたコンテンツの集合,  $C$  は推薦対象となるコンテンツの全体集合を表す。この指標を用いることで, どれほど多くのコンテンツを推薦できているかを測ることができる。

### 4.1.2 推薦理由評価

#### (1) MEP (Mean Explainability Precision)

MEP は, XAI レコメンドによる推薦理由の説明可能性を測る指標である [12]。EP (Explainability Precision) とは, 推薦コンテンツ中の説明可能な推薦コンテンツ割合で定義される。

$$Explainability Precision@N_{reason} = \frac{|Lmatch_i|}{|L_i|} \quad (3)$$

$N_{reason}$  は各ユーザに対する推薦理由の提示数,  $Lmatch_i$  はユーザ  $i$  に提示した推薦理由と合致した推薦コンテンツの集合,  $L_i$  はユーザ  $i$  に推薦したコンテンツの集合を表す。EP の全ユーザ平均をとったものが MEP である。この指標を用いることで, どれほど根拠を持ってコンテンツを推薦できているかを測ることができる。

#### (2) MER (Mean Explainability Recall)

MER は MEP と同じく XAI レコメンドによる推薦理由の説明可能性を測る指標である [12]。ER (Explainability Recall)

とは、説明可能な全コンテンツ中の説明可能な推薦コンテンツの割合で定義される。

$$\text{Explainability Recall}@N_{reason} = \frac{|Lmatch_i|}{|Cmatch_i|} \quad (4)$$

$Cmatch_i$ はユーザ*i*に提示した推薦理由と合致する全コンテンツの集合を表す。ERの全ユーザ平均をとったものがMERである。この指標を用いることで、どれほど取りこぼし無く、推薦理由に合うコンテンツを推薦できているかを測ることができる。

### (3) 推薦理由精度

推薦理由精度は、推薦理由と合致するコンテンツを評価期間中にコンバージョンしているユーザの割合で定義する。

$$\text{Reason Accuracy}@N_{reason} = \frac{|Umatch_{cv}|}{|U|} \quad (5)$$

$Umatch_i$ はユーザ*i*に提示した推薦理由と合致するコンテンツをコンバージョンしたユーザの集合を表す。この指標を用いることで、推薦理由として提示した情報が各ユーザの興味を引くかどうかを評価期間中のコンバージョンから推測することができる。

### (4) カバレッジ

カバレッジは、全推薦理由候補中の各ユーザに一度でも提示された推薦理由の割合で定義する。

$$\text{Coverage}@N_{reason} = \frac{|R_{show}|}{|R|} \quad (6)$$

$R_{show}$ は各ユーザに一度でも提示された推薦理由の集合、 $R$ は提示対象となる全推薦理由候補の集合を表す。この指標を用いることで、どれほど多くの推薦理由をユーザに提示できているかを測ることができる。

## 4.2 評価条件

XAIレコメンドモデルの評価にあたり、株式会社NTTドコモが提供している動画配信サービスのデータを評価対象データとする。本データには各ユーザの視聴履歴、コンテンツのメタ情報が付与されており、モデルの学習、評価に用いる。視聴履歴データ期間は2022年1月17日～2022年1月29日の13日間、評価対象ユーザ数は10,000名、推薦対象コンテンツ数は4,513作品である。式(1)や式(5)のような精度評価を行うため、モデル学習期間を2022年1月17日～2022年1月27日、評価期間を2022年1月28日～2022年1月29日と分割する。学習期間の履歴で学習したモデルを用いて各ユーザへの推薦コンテンツや推薦理由を決定し、そのうち評価期間にコンバージョンしたコンテンツや推薦理由の割合で評価を行う。推薦理由の候補となるのは本コンテンツメタ情報中の「ジャンル」29個、「出演者」100個とする。

2章で述べたように、本研究ではXAIレコメンドモデルであるAMCFのオフライン評価を行う。また、AMCFの有用性を確かめるため、AMCFに加え、LightGBM[13]にSHAP[14]を適用したモデル、ランダム提示と比較評価を行った。LightGBMは機械学習のコンペティションでも多用される高精度な機械学習モデルであり、SHAPは最も有名なXAI技術のひとつである。今回のXAIレコメンドタスクにおいても高い性能を達成できると推測し採用した。また、ランダム提示は評価値におけるベースラインとして用いた。推薦コンテンツ数 $N_{recommend}$ と推薦理由提示数 $N_{reason}$ は各モデルで統一し、それぞれ10個、3個とした。

### (1) AMCF

ユーザ視聴履歴より、各ユーザの各コンテンツに対する視聴回数を計算しモデルに入力する。また、ユーザ嗜好を推定するため、入力する視聴履歴には各コンテンツのメタ情報をMulti-hot Vectorの形式で紐付ける。それにより、各ユーザへの推薦コンテンツ、ユーザ毎に推定される「入力されたコンテンツメタ情報のうちのメタ情報を好むか」といったユーザ嗜好を得ることができる。モデルの出力した推薦コンテンツ、ユーザ嗜好を推薦理由として用いて評価を行う。

### (2) LightGBM+SHAP

ユーザID、コンテンツIDをそれぞれインデックス化したもの、各コンテンツのメタ情報をMulti-hot Vector形式にしたものをLightGBMへの入力として用いる。また、AMCFと同様にユーザ視聴履歴より各ユーザの各コンテンツに対する視聴回数を計算し正解ラベルとして使用し学習、回帰問題として推薦コンテンツを決定する。推薦理由の決定にはSHAP (SHapley Additive exPlanations) 値を用いる。ユーザ毎に推薦コンテンツ決定に寄与したコンテンツメタ情報を決定するため、以下の計算を行った。

1. 各ユーザの各コンテンツに対するSHAP値を計算
2. SHAP値を絶対値化
3. ユーザ毎に2で絶対値化した各コンテンツに対するSHAP値を平均化

以上の推薦コンテンツ、推薦理由を用いて評価を行う。

### (3) ランダム

推薦コンテンツ、推薦理由共にランダムに選出する。推薦コンテンツは4.2項で示した全推薦対象コンテンツ4,513作品、推薦理由には全コンテンツメタ情報中の「ジャンル」29個、「出演者」100個の中からユーザ毎にランダムに選出し評価を行う。

表 1 オフライン評価結果

モデル	レコメンドモデル評価		推薦理由評価			
	推薦コンテンツ	カタログ	MEP	MER	推薦理由	カバレッジ
	精度	カバレッジ			精度	
AMCF	<b>0.2146</b>	<b>0.1994</b>	0.6052	0.0048	<b>0.6648</b>	<b>0.5891</b>
LightGBM+SHAP	0.0327	0.0348	<b>0.7987</b>	<b>0.0080</b>	0.1490	0.0853
ランダム	0.0053	1.0000	0.1243	0.0021	0.3681	1.0000

### 4.3 評価結果と考察

表 1 では、各モデルにおいて、レコメンドモデル評価、推薦理由評価に分けてオフライン評価結果を示している。以降よりレコメンドモデル評価、推薦理由評価について結果と考察を述べ、最後に総合的な評価を述べる。

#### 4.3.1 レコメンドモデル評価

推薦コンテンツ精度では 3 モデル中 AMCF が最も高い値を得られ、カタログカバレッジではランダムを除き AMCF が最も高い値となった。

この結果より、AMCF が各ユーザに幅広いコンテンツを推薦でき、かつ各ユーザの興味を惹くコンテンツを推薦できていることが示された。AMCF は純粋なレコメンドモデルとしても精度が高く従来より多用されている協調フィルタリングベースの XAI レコメンドモデルであり、レコメンドモデル評価においてもこのように一定以上の評価結果が得られたものだと考える。

一方で、LightGBM+SHAP では推薦コンテンツ精度の値が AMCF と比較しても大幅に低く、推薦コンテンツをパーソナライズできていないことが分かる。さらに、カタログカバレッジの値が 0.0348 と、全ユーザにほぼ同じコンテンツを推薦してしまっていることが分かる。LightGBM は他分野では多用されている機械学習アルゴリズムだが、このように今回の評価条件では高い精度が得られなかった。原因としてデータの入力方法が考えられる。今回は XAI レコメンドモデルの検証のため、コンテンツメタ情報ひとつひとつを特徴量とし、各特徴量に対する SHAP 値を推薦理由として用いていた。しかし、本来はコンテンツメタ情報のようなカテゴリ変数は整数でエンコード、インデックス化すべきである。入力形式の変更やチューニングによってレコメンドモデルの評価としては改善する可能性がある。

#### 4.3.2 推薦理由評価

MEP, MER では LightGBM+SHAP が最も高い値、推薦理由精度では AMCF が最も高い値、カバレッジではランダムを除き AMCF が最も高い値となっている。MEP, MER について LightGBM+SHAP が AMCF を上回った原因について述べる。LightGBM+SHAP が出力した推薦理由はコンテンツメタ情報「ジャンル」、「出演者」のうち、「ジャンル」

に偏っていたことが確認された。4.2 節で述べたように、推薦理由の候補となる「ジャンル」は 29 個、「出演者」は 100 個と、候補数に差がある。MEP と MER は図 4 のように表され、推薦コンテンツと推薦理由合致コンテンツの和集合が多くなると 2 つの指標の値は高くなる。そのため、「ジャンル」に偏った結果、図 4 中の C に該当しやすくなり、MEP, MER の値が高くなったと考えられる。

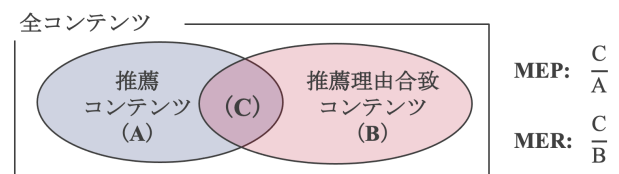


図 4 MEP, MER を表すベン図

推薦理由精度、カバレッジに関しては AMCF が LightGBM+SHAP を上回り、ユーザの興味を惹く推薦理由を幅広く提示できていることが分かる。LightGBM+SHAP に関しては 4.3.1 項や本項で前述した通り、全ユーザにほぼ同じコンテンツ、推薦理由を提示していたため、推薦理由精度、カバレッジ共に低くなっていると考えられる。

#### 4.3.3 総合評価

AMCF は MEP, MER を除き XAI レコメンドモデルの要件となる指標に関して他モデルを上回る結果となった。各ユーザの興味を惹くコンテンツ、推薦理由を幅広く提示できていることから、今回の比較においては AMCF が有用であることがわかった。さらに、2.4 節で述べた通り、AMCF は入力データに商品レビュー文章などの収集が難しいデータは不要であり、動画配信サービスのみでなく、他ドメインでも汎用的に利用可能なモデルである。以上より、精度面、汎用性を考慮しても実用的であると考えられる。

## 5. まとめと今後の課題

本研究では、レコメンドシステムがユーザにコンテンツを推薦する際に提示する推薦文言に着目し、複数モデルのオフライン評価を行った。評価は実際の動画配信サービスの視聴ログを用いてレコメンドモデル評価と推薦理由評価に分けて実施した。評価指標にはレコメンドモデル評価と

して推薦コンテンツ精度とカタログカバレッジ、推薦理由評価には MEP, MER, 推薦理由精度, カバレッジを用いることで総合的に評価を行った。評価の結果, レコメンドモデル評価に関しては推薦コンテンツ精度, カatalogカバレッジの値で AMCF が他モデルを上回り, 推薦理由評価に関しては推薦理由精度, カバレッジの値で他モデルを上回った。これにより AMCF が各ユーザの興味を惹くコンテンツ, 推薦理由を幅広く提示できていることがわかった。また, AMCF は他の XAI レコメンドモデルで使われている商品レビュー文章など特異な入力データは必要なく, サービスドメインが異なっても汎用的に利用可能なモデルという点で実用的である。

本研究ではコンテンツメタ情報としてユニーク数の異なる「ジャンル」と「出演者」をどちらも同じモデルに入力し学習, 評価を行ったが, 適切ではない可能性がある。

「ジャンル」, 「出演者」をそれぞれ別モデルで学習させることで各ユーザにとってより興味を惹く推薦理由を提示できると考えられ, 今後の課題としたい。また, 本研究では評価をオフラインで実施しており, 実際のユーザの反応を考慮した評価になっていない。XAI レコメンドモデルのみでなく通常のレコメンドモデルにおいても言えることだが, 実際にユーザに推薦コンテンツや推薦理由を提示し CTR などを評価することが最も適切な評価方法である。提示した推薦コンテンツや推薦理由が実際にユーザの興味を惹いたのかを示すため, オンラインでの評価についても今後の課題としたい。

## 参考文献

- [1] Y. Zhang, X. Chen, "Explainable Recommendation: A Survey and New Perspectives," *Foundations and Trends® in Information Retrieval*: Vol. 14, No. 1, pp 1-101, 2018.
- [2] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu and S. Ma, "Explicit Factor Models for Explainable Recommendation Based on Phrase-Level Sentiment Analysis," In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (SIGIR '14)*. Association for Computing Machinery, New York, NY, USA, 83-92, 2014.
- [3] B. Abdollahi and O. Nasraoui, "Explainable Matrix Factorization for Collaborative Filtering," In *Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 5-6, 2016.
- [4] S. Seo, J. Huang, H. Yang and Y. Liu, "Interpretable Convolutional Neural Networks with Dual Local and Global Attention for Review Rating Prediction," In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17)*. Association for Computing Machinery, New York, NY, USA, 297-305, 2017.
- [5] S. Chang, F. M. Harper and L. G. Terveen, "Crowd-Based Personalized Natural Language Explanations for Recommendations," In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 175-182, 2016.
- [6] X. He, T. Chen, M.-Y. Kan and X. Chen, "TriRank: Review-Aware Explainable Recommendation by Modeling Aspects," In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM '15)*. Association for Computing Machinery, New York, NY, USA, 1661-1670, 2015.
- [7] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. V. Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston and D. Sampath, "The YouTube video recommendation system," In *Proceedings of the fourth ACM conference on Recommender systems (RecSys '10)*. Association for Computing Machinery, New York, NY, USA, 293-296, 2010.
- [8] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," *Information Fusion*, Volume 58, Pages 82-115, 2020.
- [9] M. F. Dacrema, P. Cremonesi and D. Jannach, "Are we really making much progress? A worrying analysis of recent neural recommendation approaches," In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19)*. Association for Computing Machinery, New York, NY, USA, 101-109, 2019.
- [10] D. Pan, X. Li, X. Li and D. Zhu, "Explainable Recommendation via Interpretable Feature Mapping and Evaluation of Explainability," In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, {IJCAI-20}*, 2690-2696, 2020.
- [11] J. L. Herlocker, J. A. Konstan, L. G. Terveen and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.* 22, 1 (January 2004), 5-53, 2004.
- [12] B. Abdollahi and O. Nasraoui, "Using Explainability for Constrained Matrix Factorization," In *Proceedings of the*

Eleventh ACM Conference on Recommender Systems (RecSys '17). Association for Computing Machinery, New York, NY, USA, 79–83, 2017.

- [13] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye and T.-Y. Liu, "LightGBM: a highly efficient gradient boosting decision tree," In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 3149–3157, 2017.
- [14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 4768–4777, 2017.