

# レコメンドモデルにおける CTR と相関の高いオフライン検証指標の提案

伊藤拓<sup>†1</sup> 田中茂樹<sup>†1</sup> 落合桂一<sup>†1</sup> 深澤佑介<sup>†1</sup>

**概要**：近年、Web サービスのコンテンツを推薦するためのレコメンドモデルの研究が盛んに行われている。レコメンドモデルはオンラインでの CTR を向上させることが目的であるが、研究の多くは簡単のためオフライン検証の精度指標である MAP@k, AUC SCORE を用いて評価しているものが多い。オンライン指標とオフライン指標は必ずしも相関があるわけではないので、オフライン検証で精度の高いモデルの CTR が実際に高いとは限らない。本論文では、CTR と相関の高いオフライン指標として、複数の指標(レコメンド結果の新規性、多様性、意外性、およびレコメンド対象のユーザの視聴傾向の多様性)を提案する。CTR と提案指標の相関を評価するために、動画配信サービスのある期間の CTR を提案指標から予測する検証を行った。MAP@k, AUC SCORE のみを用いて CTR を予測した場合の予測誤差(MSE)は 0.6 であったが、上記提案指標を考慮した予測手法では CTR の MSE が 0.2 に減少することを確認した。

**キーワード**：Web サービス, レコメンド, CTR, 唯一性, ユーザ多様性

## 1. はじめに

Web 上のコンテンツ配信サービスの多くは、ユーザへのコンテンツレコメンドシステムを有している。レコメンドシステムは、ユーザが興味を惹かれるコンテンツを複数提示することによって、高い CTR を獲得することが求められている。レコメンドシステムの CTR 向上のためのモデル評価は簡単のため、実サービスの運用で蓄積されたログをローカル環境にダウンロードし、オフラインで行われることが多い。例えば Lian ら[1]は、レコメンドモデルの評価方法として、広告クリックのデータセットを用いており、ある広告がクリックされる確率をモデルによって求め、実際のクリック率との差異の大きさをモデルを評価している。また Guo ら[2]は、App Store のアプリケーションのクリックログデータセットを用いて、ゲームアプリのクリック予測を行っている。

上記研究では、ユーザがどのコンテンツをクリックするかを予測するタスクを考え、そのタスクの予測精度をモデルの評価指標としている。レコメンドシステムにおいては、システムがユーザにコンテンツを提示し、ユーザが提示されたコンテンツに興味を持ち、その結果としてコンテンツをクリックする。一方で、オフライン検証での予測タスクは、ユーザが自発的にクリックしたコンテンツを予測している。そのため、オフライン検証で高い予測精度を出しているモデルが、オンラインで高い CTR を得られるとは限らない。

オフライン検証での予測精度とオンラインの CTR が一致しない原因としては、予測精度が非常に高いレコメンドシステムが提示するコンテンツは、ユーザにとって退屈なものになる可能性があることがあげられる。例えば、ヒーローものを好んで視聴するユーザに、システムが同じくヒーローものを推薦した状況を考える。この場合、ユーザの

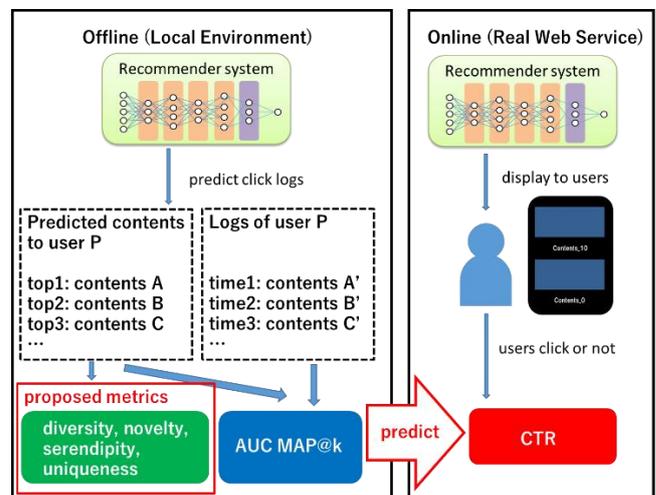


図1 本研究の概念図。提案するオフライン検証指標により、CTR に近い評価を行うことを目指す。

Figure 1 Conceptual Scheme of this Research. We Aim to Make Offline Metrics Close to CTR.

視聴傾向を捉えることはできているため予測精度は高くなるが、推薦されたユーザにとっては当たり前の結果となるため、興味を惹きづらい。Cheng ら[3]は、モデルの評価にオンラインの CTR とオフラインでの AUC を用いてモデルを評価しているが、CTR と AUC のモデルの優劣が異なる結果となっている。著者はその理由について、オフライン検証ではユーザへの提示コンテンツが固定されているために、本来推薦していればクリックされたコンテンツの評価ができないのに対し、オンライン検証では随時更新できるためであると述べている。

オフライン検証の予測精度と CTR との乖離の問題を解決するため、レコメンドシステムの多様性、新規性、意外性を評価しようという研究が存在する[4]。しかしながら、こうした指標を基に CTR を予測することを目指した研究

<sup>†1</sup> 株式会社 NTT ドコモ  
NTT docomo Inc.

はほとんど存在しない。

本研究では、CTR と相関の高いオフライン検証の評価指標として、上記3指標のほか、レコメンド結果の唯一性、レコメンド対象ユーザの多様性を提案する。本提案により、オフライン検証で CTR に近い評価を行えるようになる。そのため、モデルのチューニングを実サービスに載せなくても行えるようになり、時間的コストの低減が期待できる。

本論文の貢献は以下の通りである。

- CTR と相関の高いオフライン評価指標として、レコメンドコンテンツの多様性、新規性、意外性の他に、レコメンドコンテンツがユーザによって出し分けできているかを示す指標である唯一性を提案した。

- 上記に加え、推薦されるユーザの特性に着目し、ユーザの視聴コンテンツの多様性をオフライン評価指標として提案した。

- 提案指標である唯一性およびユーザ多様性をもとに、オフライン評価指標から CTR を予測する問題を解き、提案指標が CTR と相関の高いことを確認した。

本稿の構成は以下の通りである。次章にてレコメンドシステムの評価に関する研究について概観する。次に、3章でオフライン検証の評価指標を提案し、4章では提案指標とCTRの相関の評価を行い、提案指標の妥当性を確認する。最後に5章にて本研究のまとめおよび今後の課題について述べる。

## 2. 関連研究

### 2.1 オンライン/オフライン評価に関する研究

Yi ら[5]は、オフライン評価指標である AUC と RIG について、複数の状況を仮定して CTR との差異が大きくなるパターンについての分析を行った。彼らは論文中で、クリックと非クリックの割合に大きく差がある場合に、適切でないモデルを用いても AUC が高くなる傾向があることを発見した。

Garcin ら[6]は、ニュースのレコメンドシステムにおいて複数のモデルでオンライン検証とオフライン検証を行い、その特性を調べた。コンテキストに合わせたレコメンドを行う Context Trees モデル[7]と、直近で最も人気のあるニュース記事を推薦する Most Popular モデルを比較した結果、オフライン検証では Most Popular モデルの精度が上回っているのに対し、オンライン検証では Context Trees モデルの精度が上回っていた。この原因について彼らは、既視感のあるコンテンツを推薦されたときにユーザが興味を惹かれないからであると述べている。

### 2.2 ユーザ経験に基づく評価に関する研究

Loepp ら[8]は、レコメンドシステムの評価にはユーザ体験の満足度を測ることが重要であると述べている。一方で、

---

### Algorithm 1: CTR Prediction Using Offline Metrics

---

**Input:** Recommended contents  $R_u$  to user  $u$  by the system, and the click history  $H_u$  of user  $u$ .

**Output:** Predicted CTR  $ctr$

//Calculate the metrics

**for** each user  $u$  in all user set  $U$  **do**

    Calculate Diversity  $D_u(R_u, H_u)$

    Novelty  $N_u(R_u, H_u)$

    Serendipity  $S_u(R_u, H_u)$

    Users Diversity  $U_u(R_u, H_u)$

Calculate Uniqueness  $Q(U_u, R_u)$

//Calculate the mean of all users' metrics

$D \leftarrow \overline{D_u}$

$N \leftarrow \overline{N_u}$

$S \leftarrow \overline{S_u}$

$U \leftarrow \overline{U_u}$

//Predict CTR

Predict  $ctr(D, N, S, U, Q)$  by XGBoost

---

ユーザ体験の満足度に関する研究[9][10][11][12]は実際に推薦されたコンテンツをユーザが消費することなく評価しているため、実サービスでの評価とは異なる場合が多いとしている。そのため、レコメンドシステムの評価はあくまで推薦コンテンツをユーザが本当に消費したかを考慮する必要があるというのがこの論文の主張である。

### 2.3 関連研究まとめ

オンライン評価とオフライン評価の差異についての問題提起は過去になされている。しかし、実際にオンライン評価に有効なオフライン評価の提案や、それに基づくオンライン評価の予測を行った研究は存在しない。

また、レコメンドシステムの評価としてユーザの主観評価を採用している研究は多いが、その評価と実サービスでの消費行動は必ずしも相関があるものではない。

## 3. 提案手法

本研究では、CTR と相関の高いオフライン検証指標について検討する。[4]で定義されたレコメンドシステムの評価指標に加え、レコメンド内容の唯一性、およびレコメンド対象となるユーザのコンテンツ消費多様性をオフライン検証の指標として加えることで、よりオンラインでの CTR に近い評価をオフラインで行うことを目指す。本提案手法のアルゴリズムを Algorithm 1 に示す。

唯一性をオフライン評価に含める理由を述べる。唯一性

が高い、すなわちユーザによってレコメンドコンテンツが正しく出し分けできているということは、コンテンツ配信側が幅広くコンテンツを配信出来ているということである。したがって、多様なコンテンツの中から、対象となるユーザに適したコンテンツをレコメンドできる可能性が高くなる。逆に、唯一性が低ければ、多様なコンテンツを配信することができず、限られた内容のコンテンツからユーザにコンテンツをレコメンドすることになり、ユーザの興味を惹くことは難しくなる。このように、コンテンツ配信の状況も含めて評価することで、ユーザの興味の惹かれやすさを予測することができると考えられる。

次に、ユーザ多様性をオフライン評価に含める理由を述べる。1章でも述べたように、オフライン検証で予測精度の高いレコメンドモデルは、実サービスにおいてユーザに退屈な推薦を行う可能性が高い。実際にユーザが新鮮味のない推薦コンテンツを提示された場合に、実際にそのコンテンツを消費するかどうかはユーザの消費傾向に依存している。つまり、ユーザ多様性が高いユーザは新鮮味のないコンテンツを推薦しても消費する可能性は低く、ユーザ多様性が低いユーザは消費する可能性が高いと言える。

### 3.1 レコメンド内容に対する評価指標

本項では、ユーザに提示したレコメンドコンテンツのリストについて、多様性、新規性、意外性、唯一性を定義する。

#### (1) 多様性(diversity)

$$diversity = \frac{1}{|U|} \sum_{u \in U} \sum_{i, j \in R_u, i < j} \frac{\sqrt{prefs(i)} \times \sqrt{prefs(j)}}{prefs(i \text{ and } j)}$$

ここで、 $prefs(i)$ はコンテンツ $i$ を消費したユーザの人数、 $prefs(i \text{ and } j)$ はコンテンツ $i$ と $j$ を両方消費したユーザの人数、 $R_u$ はユーザ $u$ に対してレコメンドしたコンテンツの集合、 $U$ はユーザの全集合を表している。

上記定義がレコメンドコンテンツの多様性を表している。本研究では、コンテンツの多様性が高いという状態を、内容の似ていないコンテンツ同士がレコメンドされている状態であると定義する。コンテンツ $i$ と $j$ が似た内容のコンテンツであれば、そのコンテンツを過去消費したユーザ層は重複が多く、逆に異なる内容のコンテンツであれば、過去消費したユーザ層は重複が少ないと考えられる。したがって、レコメンドコンテンツの多様性が高ければ、内容が似通っていないコンテンツがリストに含まれるため、消費ユーザ層は重複が少なく、 $prefs(i \text{ and } j)$ は小さくなる。すなわち、上記定義で表される diversity は値が大きくなる。

#### (2) 新規性(novelty)

$$novelty = \frac{1}{|U|} \sum_{u \in U} \sum_{i \in R_u} \frac{\log_2 \frac{|U|}{prefs(i)}}{|R_u|}$$

上記定義がレコメンドコンテンツの新規性を表している。本研究では、レコメンドコンテンツの新規性を、そのコンテンツを消費したユーザが少なさで定義している。したがって、コンテンツ $i$ の新規性が高いということは、消費したユーザが少ないということであり、 $prefs(i)$ は小さくなる。すなわち、上記定義で表される novelty は高くなる。

#### (3) 意外性(serendipity)

##### serendipity

$$= \frac{1}{|U|} \sum_{u \in U} \frac{1}{|H_u|} \sum_{h \in H_u} \sum_{i \in R_u} \frac{\sqrt{prefs(i)} \times \sqrt{prefs(h)}}{prefs(i \text{ and } h)}$$

ここで、 $H_u$ はユーザ $u$ が過去消費したコンテンツの集合を表している。

上記定義がレコメンドコンテンツの意外性を表している。本研究では、レコメンドコンテンツの意外性が高いという状態を、過去ユーザが消費したコンテンツと、レコメンドコンテンツとの内容が似通っていない状態であると定義する。多様性の定義でも述べたように、コンテンツの内容が類似していないということは、消費したユーザ層が重複していないということを意味している。そのため、ユーザが過去に消費したコンテンツとレコメンドコンテンツの内容が異なっていれば、上記定義で表される serendipity は高くなる。

#### (4) 唯一性(uniqueness)

$$uniqueness = \frac{Unique(\sum_{u \in U} \sum_{i \in R_u} i)}{\sum_{u \in U} |R_u|}$$

ここで、 $Unique(\sum_{u \in U} \sum_{i \in R_u} i)$ は全ユーザの全レコメンドコンテンツの中で、ユニークなコンテンツの数を表している。つまり、コンテンツ配信側が、幅広いコンテンツを配信し、それにより、異なるユーザに異なるレコメンドコンテンツを提示できていれば、上記定義で表される uniqueness は高くなる。

### 3.2 ユーザの消費傾向を表す指標

本項では、レコメンド対象となるユーザのコンテンツ消費傾向を表す指標を定義する。

#### (5) ユーザ多様性(users diversity)

##### users diversity

$$= \frac{1}{|U|} \sum_{u \in U} \sum_{i, j \in H_u, i < j} \frac{\sqrt{prefs(i)} \times \sqrt{prefs(j)}}{prefs(i \text{ and } j)}$$

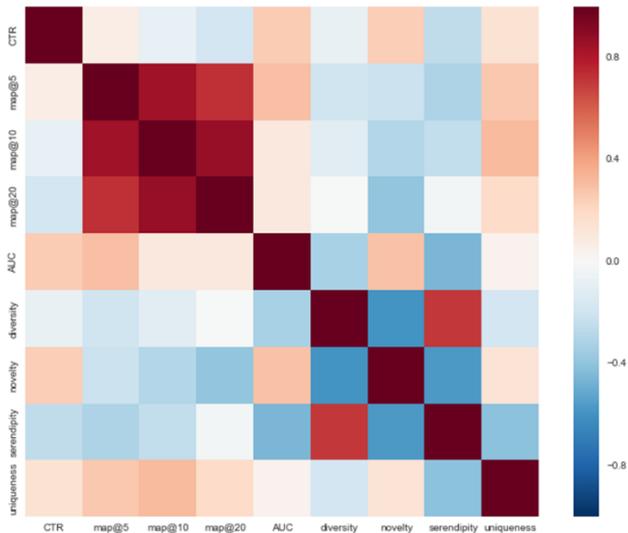


図2 CTR とオフライン検証の相関係数

Figure 2 Correlation between CTR and Offline Metrics

上記定義は、(1)の多様性とほぼ同じ定義であるが、多様性を評価する対象が「ユーザへの Recommend コンテンツ」ではなく「ユーザの消費コンテンツ」となっている。つまり、対象となるユーザが似通ったコンテンツばかりを好んで消費するユーザなのか、それとも時によって全く異なるコンテンツを消費するユーザなのかを示す指標となる。

## 4. 評価

本章では、3章で提案したオフライン評価指標について、実際に CTR との相関の高さを評価する。

### 4.1 評価対象

本研究では、株式会社 NTT ドコモが提供しているリアルタイム動画配信サービス「dTV チャンネル」を対象とする。dTV チャンネルは実際の動画視聴に加え、将来配信される予定の動画を視聴予約することができ、そのいずれかの操作を「クリック」と定義する。

本研究では、dTV チャンネルで実装されている「あなたへのおすすめ」機能における CTR と、クリックログを用いたオフライン検証結果とを比較する。

### 4.2 評価方法

2018/9/13~10/21 の 39 日間における dTV チャンネル Recommend システムの CTR 39 日分と、各日のオフライン評価結果 39 日分を用いて評価を行う。

オンライン評価、オフライン評価ともに対象となる日の前 90 日分のログを用いて Recommend モデルを学習させる。オンライン評価では、学習結果をもとに実際にアプリケーション上でユーザに推薦コンテンツを提示し、1 日の CTR

表 1 CTR 予測タスクの MSE

Table 1 MSE of CTR Prediction Task

説明変数	MSE	改善率
予測精度のみ	0.611	-
予測精度+ Recommend 内容評価指標	0.274	55.2%
予測精度+ Recommend 内容評価指標+ ユーザ多様性	0.222	63.7%

を計測する。オフライン評価では、その日のログをあとから集計し、予測精度、および 3 章で定義した各評価指標を計算する。

オフライン評価で用いる予測精度について説明する。予測精度としては、Recommend モデルの研究で多く用いられている ROC\_AUC SCORE (AUC) および Mean Average Precision@k (MAP@k) を用いる。これら 2 つの指標は、Recommend モデルが算出する 0 から 1 までのスコアと、実際のクリックの有無(クリック有:1, クリック無:0)によって計算される。AUC は、コンテンツをスコアの高い順にソートした際に、上位にクリックされたコンテンツが占める割合を表す。MAP@k は、k 個 Recommend コンテンツを提示した際に、k 個の中にクリックされたコンテンツが存在する割合を表す。

### 4.3 評価基準

本研究では、上記オンライン評価結果とオフライン評価結果をもとに、以下の 2 つの観点から相関の高さを評価する。

#### (1) 相関係数

上で述べた 39 日分の CTR と、オフライン評価指標の相関係数を計算する。3 章での提案指標と CTR の相関が高ければ、提案指標は妥当であると言える。

#### (2) CTR 予測タスクの精度

39 日分の CTR に対し、30 日の学習期間と 9 日の試験期間に分け、オフライン評価指標から CTR を予測するタスクを解く。予測には XGBoost を用いる。評価指標には予測 CTR と実測 CTR の Mean Squared Error (MSE) を用いる。予測精度のみから CTR を予測するときの精度と比べ、提案指標を用いて CTR を予測するときの精度が高ければ、提案指標は妥当であると言える。

### 4.4 評価結果

#### (1) 相関係数

図 2 に CTR とオフライン評価指標との相関係数を示す。

['map@5', 'map@10', 'map@20', 'AUC']  
MSE train : 0.000, test : 0.611



図 3a 予測 CTR と実測 CTR の推移(予測精度のみ)  
(青線=実測 CTR, 緑線=予測 CTR 以下同じ)

Figure 3a Transition of Predicted CTR and Actual CTR  
(Predicted Only by Prediction Scores)

(The blue line means actual CTR, and the green line means predicted CTR. The same applies hereafter.)

['map@5', 'AUC', 'novelty', 'serendipity', 'uniqueness', 'users\_diversity']  
MSE train : 0.000, test : 0.222



図 3c 予測 CTR と実測 CTR の推移(予測精度+レコメンド内容評価指標+ユーザ多様性)

Figure 3c Transition of Predicted CTR and Actual CTR  
(predicted by Prediction Scores, Evaluation Metrics of Recommended Contents and Users Diversity)

['map@5', 'map@10', 'map@20', 'AUC', 'diversity', 'novelty', 'serendipity', 'uniqueness']  
MSE train : 0.000, test : 0.274



図 3b 予測 CTR と実測 CTR の推移(予測精度+レコメンド内容評価指標)

Figure 3b Transition of Predicted CTR and Actual CTR  
(predicted by Prediction Scores and Evaluation Metrics of Recommended Contents)

色が濃いほど 2 指標の相関が高いことを表している。結果を見ると、AUC と CTR は弱い正の相関があるが、MAP@k と CTR にはほぼ相関がないことが分かる。また、AUC と同程度の大きさで novelty, serendipity と CTR と相関があることが分かる。この結果から、予測精度と CTR には必ずしも相関がないこと、および 3 章で提案した指標と CTR に相関があることが示された。

## (2)CTR の予測タスク

表 1 に CTR 予測タスクの精度を示す。XGBoost の説明変数として、AUC, MAP@k のみを用いて予測した場合、3.1

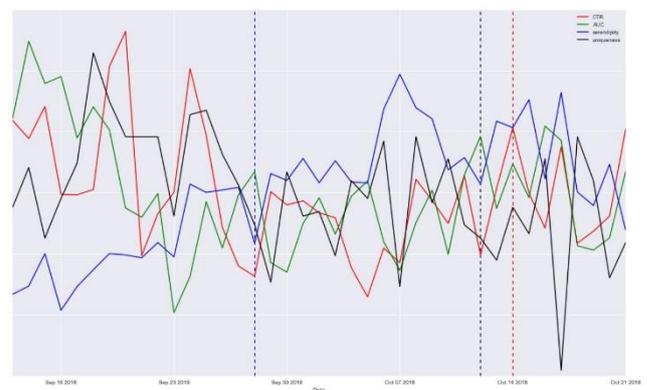


図 4 CTR, AUC, serendipity, uniqueness の推移  
(赤線=CTR, 緑線=AUC, 青線=serendipity 黒線=uniqueness)  
Figure 4 Transition of CTR, AUC, serendipity and uniqueness  
(The red line means CTR, the green line means AUC, the blue line means serendipity, and the black line means uniqueness.)

項で提案したレコメンド内容に対する評価指標を加えて予測した場合、さらに 3.2 項で提案したユーザ多様性を加えて予測した場合の 3 つの精度を比較した。

表 1 が示す通り、CTR 予測の精度は「ユーザ多様性+レコメンド内容評価指標+予測精度」>「レコメンド内容評価指標+予測精度」>「予測精度のみ」となっている。このことから、実サービスで高い CTR を出すレコメンドモデルの評価は、AUC や MAP@k といった予測精度のみで行うのは不十分であり、実際には推薦結果の多様性、新規性、意外性、唯一性、および推薦されるユーザの消費多様性も含めて評価する必要があることが示された。

次に、提案指標が有効であることの理由について考察する。図 3 は、日ごとの予測 CTR と実測 CTR を、上記 3 パターンについて載せたものである。グラフ中破線の左側は学習期間 30 日間、破線の右側は試験期間 9 日間を表して

いる。また、図 4 は 39 日間について、CTR, AUC, serendipity, uniqueness の推移をグラフで表したものである。図 4 において、赤色、青色、黒色の点線の日の結果について考察する。赤点線の日は、CTR, AUC がともに他の日と比べて高くなっている。この結果は一般的な知見とも一致し、オフラインでの予測精度の高さが実サービスでの CTR の高さとも一致している。一方で、青点線の日は、AUC は高い値を示しているが、CTR は他の日と比べて低くなっている。AUC 以外の指標を見ると、この日は serendipity も他の日と比べて低い値を示している。CTR と AUC の高さが一致しない原因としては、この日ユーザに提示された推薦コンテンツは serendipity が低い、すなわち意外性のあるレコメンドがあまりできておらず、そのためにユーザが推薦結果に興味を惹かれなかったのではないかと考えられる。同様に、黒点線の日も AUC は高いが、CTR は低くなっている。この日については、uniqueness が他の日よりも低くなっている。そのため、この日は幅広いコンテンツを配信することができず、それによりレコメンドの幅が狭まり、CTR に結び付かなかったのではないかと考えられる。

以上の考察のように、レコメンドモデルの実サービスでの CTR をオフライン検証で考える際には、クリックの予測精度だけでなく、推薦内容や推薦対象となるユーザを考慮した指標を含めて評価する必要があることが示された。

## 5. おわりに

本研究では、Web サービスのコンテンツ配信サービスにおけるレコメンドシステムの評価に着目し、オフライン検証時の評価と実サービスでの CTR に乖離がある課題を解決することを目指した。そのためにレコメンドシステムの s オフライン検証で一般的に用いられている評価指標である予測精度(AUC, MAP@k)の他に、推薦コンテンツの多様性、新規性、意外性、唯一性、および推薦対象のユーザのコンテンツ消費の多様性に着目し、オフライン検証時の評価指標に組み込むことで、より実サービスでの CTR に近い評価を行えるか検討した。提案指標の有効性を示すため、動画配信サービスにおいて 39 日間分の CTR, およびオフライン検証結果を抽出し、提案指標と CTR との相関係数を求める評価実験と、提案指標から CTR を予測するタスクの精度を検証する評価実験の 2 つを行った。評価の結果、novelty と serendipity において、CTR との相関が AUC と CTR との相関とほぼ同等の大きさであることを確認した。また、CTR 予測タスクにおいては、推薦コンテンツの内容を考慮した評価指標を説明変数に加えると、予測精度のみときと比べて MSE が 55.2%改善し、さらにユーザ多様性も考慮すると 63.7%改善することを確認した。本研究では、提案した CTR 予測手法を用いたレコメンドモデルのチューニングまでは行えていないため、今後の課題としたい。

## 参考文献

- [1] Jianxun Lian, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie and Guangzhong Sun, “xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems”, Proceedings of the 24<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp.1754-1763, 2018.
- [2] Hui Feng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li and Xiuqiang He, “DeepFM: a factorization-machine based neural network for CTR prediction”, Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp.1725-1731, 2017.
- [3] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, Hemai Sneh, “Wide & Deep Learning for Recommender Systems”, Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, pp.7-10, 2016.
- [4] Yuan Cao Zhang, Diarmuid O, Seaghdha, Daniele Quercia and Tamas Jambor, “Auralist: Introducing serendipity into music recommendation”, Proceedings of the 5th ACM international conference on Web search and data mining, pp.13-22, 2012.
- [5] Jeonghee Yi, Ye Chen, Jie Li, Swaraj Sett and Tak W. Yan, “Predictive model performance: offline and online evaluations”, Proceedings of the 19<sup>th</sup> ACM SIGKDD international conference on Knowledge discovery and data mining, 2013.
- [6] Florent Garcin, Boi Faltings, Olivier Donatsch, Ayar Alazzawi, Christophe Bruttin and Amr Huber, “Offline and online evaluation of news recommender systems at swissinfo.ch”, Proceedings of the 8<sup>th</sup> ACM Conference on Recommender systems, pp.169-176, 2014.
- [7] Florent Garcin, Christos Dimitrakakis and Boi Faltings, “Personalized news recommendation with context trees”, Proceedings of the 7<sup>th</sup> ACM conference on Recommender systems, pp.105-112, 2013.
- [8] Benedikt Loepp, Tim Donkers, Timm Kleemann and Jurgen Ziegler, “Impact of item consumption on assessment of recommendations in user studies”, Proceedings of the 12nd ACM Conference on Recommender Systems, pp.49-53, 2018.
- [9] Michael Jugovac and Dietmar Jannach, “Interacting with Recommenders – Overview and Research Directions”, Journal of ACM Transactions on Interactive Intelligent Systems, Vol.7, Issue 3, No. 10, 2017.
- [10] Joseph A Konstan and John Riedl, “Recommender systems: from algorithms to user experience”, Journal of User Modeling and User-Adapted Interaction, Vol.22, Issue 1-2, pp. 101-123, 2012.
- [11] Chen He, Denis Parra and Katrien Verbert, “Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities”, Journal of Expert Systems with Applications, Vol. 56, pp. 9-27, 2016.
- [12] Pearl Pu, Li Chen and Rong Hu, “Evaluating recommender systems from the user’s perspective: survey of the state of the art”, Journal of User Modeling and User-Adapted Interaction, Vol. 22, Issue 4-5, pp. 317-355, 2012.