動画視聴型オンラインカンファレンスにおける推薦システムの実証実験

大社　綾乃†　　大滝　啓介†　　石井　良尚†　　小出　智士†

† 株式会社豊田中央研究所　〒 480–1118 愛知県長久手市横道 41-1
E-mail: †{okoso,otaki,y-ishii,koide}@mosk.tytlabs.co.jp

あらまし　新型コロナウイルス流行の影響により，多くの学術会議がオンラインで開催されるようになった．事前に録画された発表の動画をサイト上に埋め込み，参加者が自由なタイミングで視聴する形式を「動画視聴型オンラインカンファレンス」と呼び，本研究の対象とする．動画視聴型オンラインカンファレンスでは，多くの候補から自身の興味にあう動画を効率的に見つけ出さなければならず，視聴を予定していなかったが興味が湧くような動画に出会える機会が少ないという問題点がある．このような問題点を緩和するため，推薦システムの導入を検討する．本稿では，4000 人規模の動画視聴型オンラインカンファレンスにおいて，代表的な推薦システムを実装し，その効果を A/B テストにより検証した．検証の結果，推薦システムを利用した参加者群では視聴した動画数やカテゴリ数が増加し，動画探索を促進できることが分かった．協調フィルタリングに基づき個人毎に動画を推薦した場合には，継続視聴した動画の割合が向上したことが分かり，また，意外性のある視聴体験をしたユーザが多かったことが示唆された．さらに，サイト滞在時間が短い参加者ほど推薦システムの効果が顕著に現れることが分かった．

キーワード　推薦システム，協調フィルタリング，行列分解，A/B テスト，傾向スコアマッチング

1　はじめに

新型コロナウイルス流行の影響により，多くの学術会議や展示会がオンラインで開催されるようになった．オンライン形式の利点として，時間的，経済的に負担が少ないこと [1]，環境的なコストが削減できること [2]，身体的な理由等により参加が難しかった人の機会が広がること [3] が挙げられる．一方で，人脈形成の機会が限られていること [4] や心理的疲労 [5] といった欠点も報告されている．将来的には，現地開催とオンラインの欠点を補完しあうハイブリッド開催が主流になり，オンライン形式は今後も残っていくだろうと言われている [4], [6].

開催形式に関わらず，学術会議や展示会は開催期間が限られているため，全てを訪問することは難しく，自身の興味にあったものを取捨選択しなければならない．参加者は，タイトルや概要文からどの発表や展示物を訪問するか判断することが一般的である．しかしながら，多くの研究発表や展示物から自身の興味にあうものを探し出すことは，規模によっては困難である．また現地開催の場合は，参加者が自然と目に入った発表や展示物に予期せず興味をもつといった新たな出会いも起こりやすいが，オンラインの場合は積極的に巡覧しない限り，聴講を予定していなかったものが目に入る機会が少なく，このような出会いは少ないと考えられる．すなわち，オンライン開催の場合では，多くの候補から自身の興味にあうものを効率的に見つけ出すこと，そして，予期していなかったが興味が湧くようなものに出会える機会を増やすこと，が望まれる．これらの要望を満たすため，推薦システム [7] の活用が有望であると考えられる．

推薦システムとは，大量の候補から価値のあるものを提示したり，予期せず興味が湧くものとの出会いを促進させ，ユーザの意思決定を支援する仕組みである．産業界では，推薦システムはユーザ体験の向上と販売を促進させるための重要なツールとして，商品推薦 [8] や動画推薦 [9]，ニュース推薦 [10] など，様々なサービス [11] で利用されている．一方で，学術会議や展示会を対象としたものは少ない．例えば，現地開催された展示会における展示ブースの推薦について，利用者の抵抗感を調査した研究がある [12]．しかしながら，オンラインで開催される学術会議や展示会へ推薦システムを適用させ，その効果を検証した事例は，我々の知る限り存在しない．

そこで本研究では，オンラインの学術会議において推薦システムがもたらす効果について検証する．事前に録画された発表の動画を web サイト上に埋め込み，参加者が期間中にいつでも視聴できるような形式を「動画視聴型オンラインカンファレンス」と呼ぶことにし，本研究の対象とする．以前我々は，実際の動画視聴型オンラインカンファレンスの視聴履歴データを用いたオフライン評価を通して，協調フィルタリング手法により参加者の動画視聴履歴を予測できることを示した [13]．推薦システムにより参加者の興味にあった動画を提示できることが示唆された一方で，予期せず興味の湧く動画と出会えるかどうかの評価やオンライン評価は今後の課題であった．

本稿では，動画視聴型オンラインカンファレンスにおいて推薦システムがもたらす効果を「興味との適合」と「予期せぬ発見」の観点で評価する．前者は，多くの研究発表から自身の興味にあったものを見つけられることに相当し，後者は，視聴を予定していなかったものの興味が湧いた研究発表との出会いがあったかに相当する．本研究の貢献は以下の通りである．

（1）　4000 人以上が参加する動画視聴型オンラインカンファレンスに推薦システムを導入し，A/B テスト [14] により推薦システムの効果を検証した．我々の知る限り，実際の動画視聴型オンラインカンファレンスを対象に推薦システムの効果を検

証した例は，本研究が初めてである．

（2） 検証の結果，推薦システムの導入により動画視聴数や視聴カテゴリ数が増え，動画探索を促進できることが分かった．協調フィルタリングに基づく推薦を提示した場合には，継続視聴した動画の割合が向上したことが分かった．また，アソシエーション分析により，予期せぬ発見に繋がる視聴体験をした参加者が多いことが示唆された．さらに，サイト滞在時間が短い参加者に推薦システムが特に効果的であることが分かった．

本稿の構成は次の通りである． 2 章で関連研究を述べ， 3 章で実験対象となるオンラインカンファレンスの概要について述べる． 4 章では本実験に用いた推薦手法について説明し， 5 章で実験設定および結果について述べる． 6 章で本稿をまとめ，今後の課題について論ずる．

2 関 連 研 究

本章では，オンラインカンファレンスに関する先行研究と代表的な推薦システムのアプローチについて述べる．

2.1 オンラインカンファレンス

本節では，オンライン MICE [15] に関する先行研究について述べる．MICE とは，Meeting（会議），Incentive travel（研修旅行），Convention/Conference（学会），Exhibition/Event（展示会）の頭文字からなる造語であり，経済効果の大きい重要な産業分野である．新型コロナウイルス流行の影響により，オンラインで開催される MICE が急速に普及したため，関連する研究も増加している．中でも，オンライン会議に関する先行研究は，教育 [16]，医療 [17]，ビジネス [18] など幅広い分野に及ぶ．オンライン会議の参加者の多くが精神的・肉体的に疲労し [19]，参加者の認知的な参加意欲が対面の場合よりも低いこと [20] が指摘されている．

一方で，オンラインカンファレンスに関する主な先行研究は，有効なツールの比較調査 [21], [22]，開催時のコツ [23]，利点および欠点 [1], [4] に関するものが多くを占める．先行研究 [24] では，オンラインカンファレンスへの参加動機と満足度について調査し，プログラムの内容やネットワーキングに加えて，コンテンツへの関与が満足度に有意な影響を及ぼすと述べている．すなわち，動画視聴型オンラインカンファレンスにおいて，動画視聴数や視聴時間が長いほど，総合的な満足度が高いと言えるだろう．オンラインカンファレンスにおいて，参加者の嗜好にあったものを提示することの重要性について先行研究 [15] で述べられているものの，実際に推薦システムを適用し，その効果を検証した事例は，我々の知る限り存在しない．

2.2 代表的な推薦手法

推薦を受けるユーザ（本節では対象ユーザと記す）に対してアイテムを推薦する方法は，個人化手法と非個人化手法に大別される．個人化手法では対象ユーザの嗜好を推定し，対象ユーザごとに異なる推薦結果を提示する．一方で非個人化手法では，売上に基づく人気アイテムやサービス提供者がお薦めしたいアイテムなど，全対象ユーザに対して同じ推薦結果を提示する．

対象ユーザの嗜好を推定する古典的な手法は，内容ベースフィルタリング [25] と協調フィルタリング [26] に大別される．内容ベースフィルタリングではアイテムの特徴ベクトルと対象ユーザの嗜好を比較し，対象ユーザが好むと予想されるアイテムを推薦する．アイテムと対象ユーザのデータのみを用いるため，推薦システムを利用するユーザ数が少なくても利用可能であるが，類似したアイテムが推薦されやすいといった欠点もある．協調フィルタリングではアイテムへの評価が類似しているユーザは嗜好が類似しているという直感に基づき，他のユーザの評価情報を用いてアイテムを推薦する．他のユーザ情報を用いることで，対象ユーザ自身が知らない特徴を持つアイテムを知ることができるが，他のユーザによる評価が十分でないアイテムの推薦が難しいというコールドスタート問題がある．

本稿では個人化手法と非個人化手法の両方を実装した．対象ユーザの「興味との適合」に加えて「予期せぬ発見」の効果も検証するため，個人化手法では，対象ユーザ自身が知らない特徴を持つアイテムを推薦可能な協調フィルタリングを用い，非個人化手法では人気ランキングを用いた．

3 REX’22 の概要

本章では，本稿で対象とした動画視聴型オンラインカンファレンスについて概要を述べる．以降では，このカンファレンスを “Research EXhibition”，略して REX と表記する．REX は一般には非公開であり，弊社のグループ会社向けに研究発表を行うことを目的として毎年開催される．新型コロナウイルス流行の影響により，2020 年以降 REX はオンラインで開催された．本稿では 2022 年 5 月に開催された REX’22 を扱う．

REX’22 では，67 件の研究発表が実施された．各研究発表につき約 5 分程度の発表動画があり，7 つのカテゴリのうちのいずれか 1 つに含まれる．以降では，個別の研究発表を「テーマ」，発表動画を単に「動画」，カンファレンスの参加者を「ユーザ」と表記する．図 1 のように，各テーマは異なるページに掲載され，動画再生ボタン，イイねボタン，問い合わせボタン，掲載テーマへのアンケートが表示されている．また，同じカテゴリに属する別のテーマへ遷移するボタンと，本稿で実装した「(ユーザへ）お薦めのテーマ」へ遷移するボタンもある．このテーマページへユーザが到達する方法は，以下の 4 種類である．

- トップページ → カテゴリページ → テーマページ，
- テーマページ → テーマページ，
- トップページ → 全テーマ一覧ページ → テーマページ，
- トップページ → 運営者お薦めテーマ一覧ページ → テーマページ．

REX’22 は 2 週間開催され，4,000 人以上が参加した．期間中であればユーザはいつでもサイトの閲覧，動画の視聴，アンケート回答等のインタラクションが可能である．なおユーザには REX’22 の web サイトへ最初に訪問する際，研究利用するために個人を特定できない形でイベントログを取得することについて同意を求めた．本稿では同意を得られたユーザのみのデータを用いた．また，本研究の実験プロトコルは所属機関の倫理

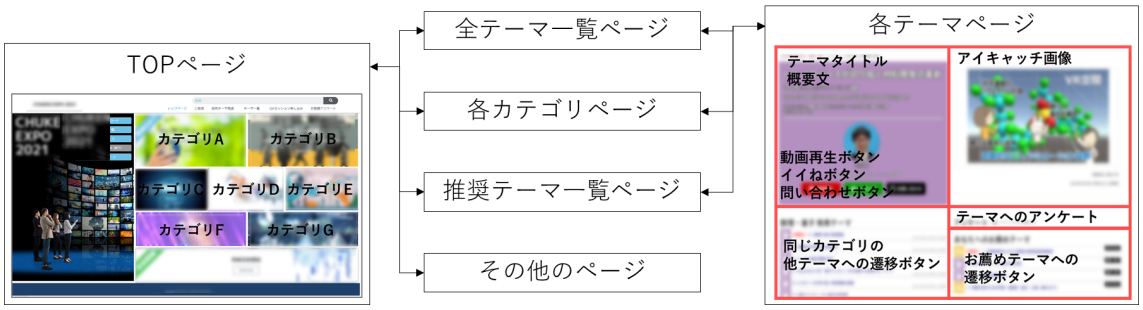


図 1: web サイトのページ構成

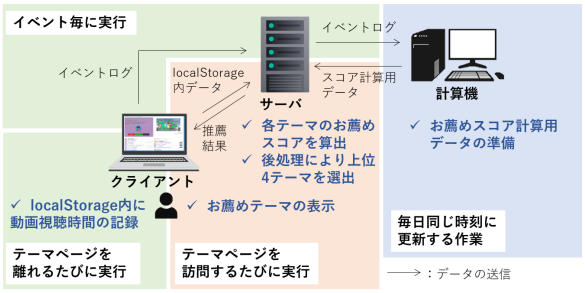


図 2: 推薦システムの処理の流れ

委員会にて承認を得ている（承認番号：21B-09）.

4 推薦システムの構築

本章では，推薦結果を提示する処理とユーザにテーマを推薦するために用いた手法について述べる．なお，各テーマにつき動画は 1 つであるため，テーマを推薦することは動画を推薦することと同じである．

4.1 処理の流れ

図 2 に示すように，ユーザがテーマページに訪問する度に，サーバ内で各テーマの嗜好スコアを計算し，後処理を施した後に嗜好スコアの高い上位 4 件を選び，テーマページに表示した．嗜好スコアの計算には，各ユーザがテーマページを訪問した時点での動画視聴履歴と蓄積された行動履歴に基づいて計算されたデータを用いた．動画視聴履歴とは，これまでに視聴開始した動画の実際の視聴時間を表し，ユーザがテーマページを離れる度に LocalStorage に記録した．なお，動画を視聴しなかった場合は視聴時間を 0 秒と記録し，同じ動画を複数回視聴した場合は視聴時間を累加した．行動履歴とは，動画視聴履歴と異なり，テーマページの訪問や動画視聴開始ボタンのクリックなど，イベントログを意味する．行動履歴から，各テーマに対応する動画の視聴時間を算出 [1] し，嗜好スコアの計算に用いた．嗜好スコアの計算の詳細は 4.2 節で述べる．嗜好スコアの計算に用いるデータは，毎日 18 時から 19 時の間に学習および更新した．後処理では，現在訪問中および訪問済みのテーマに対して嗜好スコアの割引処理を行い，訪問中および訪問済みテーマが再度推薦されないようにした．

1：データの前処理方法および算出方法は付録 1 を参照.

4.2 比較手法

本節では，ユーザにテーマを推薦するために用いた嗜好スコアの計算方法について述べる．本稿では，非個人化推薦である人気ランキング法と，協調フィルタリングに基づく個人化推薦を行うために，行列分解を用いる手法を実装した．どちらも古典的な手法であるが，その効果の高さから多くのサービスで現在も利用されている．両手法ともユーザの視聴履歴を用いるため，REX'22 の初日はデータ収集日とし，推薦システムは適用しなかった．2 日目以降は，前日までの全ての視聴履歴を用いて推薦システムを適用した．以下では，本実験で実装した推薦手法の詳細について述べる．

4.2.1 人気ランキング法

全動画からなる人気ランキングに基づき，全ユーザに対して同一の推薦結果を表示する手法を実装した．前日までの全ユーザの視聴数をその最大値で正規化した値を人気スコアとし，人気スコアが大きい順にランキングを作成した．動画の視聴判定は，90 秒以上視聴していたか否かで決定した．本手法では，この人気スコアをそのまま嗜好スコアとして用いた．

4.2.2 NMF ベース手法

行列分解は協調フィルタリング [26] の代表的な手法である．行列分解では，アイテムに対するユーザの評価値で構成された評価値行列を，アイテム埋め込み行列とユーザ埋め込み行列へ低ランク近似分解し，アイテム埋め込み行列とユーザ埋め込み行列の内積から未評価のアイテムの評価値を予測する．我々は，行列分解の中でも非負値の要素のみからなる低ランク行列に分解する手法である，非負値行列因子分解（Nonnegative Matrix Factorization; NMF）[27] を実装した．

要素 (i, j) がユーザ i のアイテム（テーマ/動画）j への評価値からなる $m \times n$ の評価値行列 $\mathbf{R}$ を，$m \times k$ のユーザ行列 $\mathbf{U}$ と $n \times k$ のアイテム行列 $\mathbf{V}$ を用いて以下のように近似する．

$$\mathbf{R} \approx \mathbf{U}\mathbf{V}^{\top} =: \hat{\mathbf{R}} \quad \text{s.t. } \mathbf{U}, \mathbf{V} \geq 0 \tag{1}$$

ここで $\top$ は転置を，$\hat{\mathbf{R}}$ は予測評価値行列を表し，$k \ll \min(m, n)$ である．本稿では，評価値行列 $\mathbf{R}$ の要素を，ユーザ i が動画 j を視聴した場合に $r_{ij} = 1$，視聴しなかった場合に $r_{ij} = 0$ とし，視聴時間が 90 秒以上か否かで視聴有無を判定した．交互最小二乗法や勾配降下法を用いて，式 (2) で定義される損失を最小化することで，$\mathbf{U}$ と $\mathbf{V}$ を求める．

$$\sum_{(i,j)\in\Omega} (r_{ij} - \mathbf{u}_i^\top \mathbf{v}_j)^2 \quad (2)$$

ここで，Ω は評価済みのデータ集合，$\mathbf{u}_i$ は $\mathbf{U}$ の第 i 行ベクトルであり，$\mathbf{v}_j$ は $\mathbf{V}^\top$ の第 j 列ベクトルを表す．$\mathbf{u}_i$ と $\mathbf{v}_j$ は，それぞれユーザ i の埋め込みベクトルとアイテム j の埋め込みベクトルを意味する．ユーザ i のアイテム j への予測評価値 $\hat{r}_{ij}$ は，$\hat{r}_{ij} = \mathbf{u}_i^\top \mathbf{v}_j$ で求めることができる．

しかしながら，評価値行列 $\mathbf{R}$ をリアルタイムで更新し，その都度，式 (1) で定義される近似分解をすることは困難である．そこで，前日までに収集されたデータで作成した評価値行列 $\mathbf{R}$ を近似分解することで $\mathbf{V}$ を事前に求めておき，推薦対象ユーザ x の埋め込みベクトルは現在の視聴履歴 $\hat{\mathbf{h}}_x$ とアイテム行列 $\mathbf{V}$ から推定した $\hat{\mathbf{u}}_x$ を用いる．具体的には，$\hat{\mathbf{u}}_x = (\mathbf{V}^\top\mathbf{V})^{-1}\mathbf{V}^\top\hat{\mathbf{h}}_x$ で $\hat{\mathbf{u}}_x$ を求め，ユーザ x のアイテム j に対する予測評価値 $\hat{r}_{xj} = \hat{\mathbf{u}}_x^\top \mathbf{v}_j$ を計算する．この予測評価値を嗜好スコアとして用いた．なお，前日までに収集されたデータで作成した評価値行列 $\mathbf{R}$ に，ユーザ x の評価情報が含まれなくても良い．以降では，本手法を NMF ベース手法と呼ぶ．

5 オンライン実験

本章では，動画視聴型オンラインカンファレンスにおいて推薦システムがもたらす効果を A/B テスト [14] により検証する．A/B テストは推薦システムの代表的なオンライン評価方法であり，ユーザをランダムに分割し，異なる推薦結果を提示することで，その効果やユーザの行動の違いを検証することができる．以下では，オンライン実験の設定および推薦システムの構築，実験結果と考察について論ずる．

5.1 群の設定と評価指標

推薦システムの有無によるユーザの視聴行動を比較するため，ユーザを以下の 3 群 [2]にランダムに割り付けた．

A 群　：推薦システムを導入しない群
B 群　：人気ランキング法に基づく非個人化推薦を提示する群
C 群　：NMF ベース手法に基づく個人化推薦を提示する群

また，B 群および C 群のうち推薦システムを利用したユーザの群を，それぞれ B+群および C+群と定義する．本実験では，推薦されたテーマへの遷移ボタンを一度でもクリックした場合に，推薦システムを利用したと定めた．

本稿では，推薦システムがもたらすと期待される効果として「興味との適合」および「予期せぬ発見」の観点に着目している．そのため，推薦結果の精度に加えて，興味との適合および予期せぬ発見に相当する指標について評価した．用いた評価指標を表 1 にまとめる．

興味との適合に対応する評価指標として，動画視聴開始率および継続視聴した動画の割合，動画視聴時間の割合を用いた．動画視聴開始率は，訪問したテーマのうち動画の視聴を開始した割合を表し，訪問したテーマが自身の興味に合いそうだと判断したかに相当する．継続視聴した動画の割合は，視聴を開始した動画数のうち 90 秒以上した動画数の割合を表し，その動画が実際に興味に適合していたかに相当する．また，動画視聴時間の割合は各動画の実時間に対して実際にどの程度の割合を視聴していたかを表す．

予期せぬ発見に対応する評価指標として，動画視聴数やカテゴリ数，非介入群である A 群での共起カテゴリからの乖離度合いを用いた．動画視聴数やカテゴリ数が多いほど，予期しないものに出会う可能性が高いと推測できる．しかしながら，これらの指標では，視聴した動画やカテゴリが実際に予期しなかったものであったかを判断することはできない．そこで，予期しなかった動画に出会うことを，一般には一緒に視聴される可能性が低い，すなわち共起率が低いカテゴリの動画を視聴することと仮定し，非介入群である A 群の共起カテゴリからの乖離の程度で予期しない発見の度合いを定義した．共起率は，A 群のデータのみを用いたアソシエーション分析 [28] におけるリフト値で定義した．詳細は 5.2.3 章で述べる．なお先行研究では，推薦結果意の意外性の指標として，過去の消費アイテムとの非類似度 [29] や，人気ランキングなどのプリミティブな手法による推薦結果との乖離度 [30] が用いられている．しかしながら，これらは推薦結果が予期しないものであるかを測定する指標である．我々は推薦システムの有無に関わらず，ユーザが予期せぬ発見を経験したか否かを評価することを目的としているため，共起カテゴリからの乖離度合いを用いた．簡単のため，5.2.3 章で述べる比較にのみ，この指標を用いた．

推薦結果の精度として，推薦リンクの CTR (Click-Through Rate) と CVR (ConVersion Rate) を評価指標とした．提示された推薦結果がユーザの興味を掻き立てるものであった場合は CTR の値が高くなり，さらに，実際に動画の視聴を開始した場合には CVR の値が高くなる．そのため，これらの指標は「興味との適合」にも相当することに注意されたい．

上述した上記で述べた推薦システムがもたらすと期待される効果と推薦結果の精度以外に，ユーザの積極性も評価した．積極性に対応する評価指標としては，サイト滞在時間およびイイねボタンクリック率，アンケート回答率を用い，ユーザがどの程度 REX に積極的に参加していたかを表す．

5.2 結　　果

5.2.1 推薦システムの利用率

各ユーザ群における推薦システムの利用率を表 2 に示す．各群に割り当てられたユーザ総数は，A 群で 1,340 人，B 群で 1,418 人，C 群で 1,376 人であった．推薦システムにより推薦結果が表示されたユーザ数は，B 群で 1,312 人，C 群で 1,121 人であった．B 群および C 群において一部のユーザに推薦結果が表示されなかった理由は，データ収集日とした初日のみ web サイトを訪れた両群のユーザや，動画を視聴しなかった C 群ユーザによるものである．B 群および C 群で推薦システムを利用したユーザ数，すなわち B+群および C+群はそれぞれ 292 人と 236 人であった．よって，推薦システムの利用率は，それ

2：ランダムに提示する群を導入することで，各推薦手法の効果をより明らかに検証可能となるが，実際のサービス運営においてユーザの満足度低下に繋がる恐れがあったため，本実験では導入しなかったことを付記しておく．

表 1: 評価指標

分類	評価指標	説明
興味との適合	動画視聴開始率	訪問したテーマページのうち動画視聴開始ボタンをクリックした割合
	継続視聴した動画の割合	動画視聴開始ボタンをクリックしたうち視聴時間が 90 秒以上だった割合
	動画視聴時間の割合	各動画の実時間に対する視聴した時間の割合の平均値
予期せぬ発見	共起カテゴリからの乖離度合い	リフト値が低いカテゴリの組合せを視聴したユーザ数の割合
	動画視聴数	90 秒以上視聴された動画数
	視聴カテゴリ数	90 秒以上視聴された動画が属するカテゴリ数
推薦結果の精度	推薦リンクの CTR	推薦結果が表示されていた回数に対する推薦結果のリンクをクリックした割合
（B 群，C 群のみ）	推薦リンクの CVR	推薦結果のリンクから訪問したテーマのうち動画視聴開始ボタンをクリックした割合
積極性	サイト滞在時間	web サイトの総滞在時間
	イイねクリック率	動画視聴開始ボタンをクリックしたテーマのうちイイねボタンをクリックした割合
	アンケート回答率	動画視聴開始ボタンをクリックしたテーマのうちアンケートに回答した割合

表 2: 各ユーザ群の推薦システム利用率

	A 群	B 群	C 群
ユーザ総数	1,340	1,418	1,376
推薦結果が表示されたユーザ数	0	1,312	1,121
推薦システムを利用したユーザ数 (B+/C+)	0	292	236
推薦システムの利用率	0%	22.3%	21.1%

ぞれ 22.3%，21.1%であった．

5. 2. 2　群間の単純比較

a）推薦システム導入有無の比較

推薦システムがもたらす効果を検証するため，各評価指標について A 群と B 群，A 群と C 群の結果を比較した．各評価指標の平均値とウェルチの t 検定の結果を表 3 に示す．継続視聴した動画の割合など一部の評価指標で有意差が見られたものの，ほとんどの評価指標で有意差が確認できなかった．しかしながら，推薦システムを利用したユーザが B 群および C 群で 20%強であったため，単純比較では，各群の全ユーザで評価すると推薦システムを利用しなかったユーザの結果に埋もれてしまい，推薦システムによる影響が捉えにくいと考えらえる．

そこで推薦システムを利用したユーザに着目し，A 群と B+群，A 群と C+群の比較を行った．表 3 に示した結果から，B+群と C+群は A 群よりもほぼ全ての指標で有意に向上する結果が得られた．しかしながら，推薦システムの影響を受けないと考えられるイイねクリック率およびアンケート回答率も向上している．これは選択バイアス [31] の可能性を示唆している．例えば「動画視聴数等が多いユーザが推薦システムを利用する傾向があったこと」と「推薦システムを利用したユーザが動画視聴数等が向上したこと」が区別できないことを意味する．このようなバイアスを可能な限り緩和するため，傾向スコアマッチング [32] を実施して比較した．比較結果は 5. 2. 3 節で述べる．

b）推薦手法間の比較

推薦手法による効果の違いを検証するため，B+群と C+群を比較した結果について論ずる．各評価指標の平均値の結果を表 3 に示す．B+群と C+群での比較検定の結果，動画視聴開始率（$p < 0.01$）および継続視聴した動画の割合（$p < 0.01$），動画視聴時間の割合（$p < 0.05$）のみ統計的な有意差が認められた．これらの指標はいずれも「興味との適合」に相当し，NMF ベース手法の方がユーザの興味をよく捉えられることを表している．ただし，NMF ベース手法は個人化推薦であるため，妥当な結果である．一方で，予期せぬ発見や積極性については統計的に有意な差はみられなかった．推薦結果の精度は，表 3 に示す通り，CTR は両群で有意差はなく，CVR は C+群の方が高い結果となった．すなわち，人気ランキング法による推薦結果は，遷移先で動画が視聴される可能性が低いことを表している．

表 3: A 群と比較した検定結果．※印の評価指標は B+群と C+群で比較．（$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$）

	A 群	B 群	C 群	B+群	C+群
ユーザ数	1340	1418	1376	292	236
動画視聴開始率	0.77	0.75	0.76	0.74	0.80**
継続視聴した動画の割合	0.77	0.75*	0.78	0.83***	0.92***
動画視聴時間の割合	0.69	0.67	0.70	0.75***	0.83***
動画視聴数	7.8	7.4	7.4	10.1***	10.4***
視聴カテゴリ数	2.7	2.7	2.7	3.7***	3.5***
サイト滞在時間（分）	67.7	62.3**	64.6	85.1***	92.5***
イイねクリック率	0.21	0.21	0.21	0.27***	0.27***
アンケート回答率	0.36	0.38	0.39**	0.52***	0.57***
推薦リンクの CTR（※）	-	0.04**	0.03	0.19	0.18
推薦リンクの CVR（※）	-	0.15	0.14	0.74	0.81**

5. 2. 3　選択バイアスを緩和した群間の比較

傾向スコアマッチング [32] を用いて，各群における前節で述べた選択バイアスを緩和した上で比較を行う．具体的には，A 群から B+/C+群と性質が類似したユーザを抽出し，B+/C+群と比較する．A 群および B+群の傾向スコアマッチングにより A 群から抽出されたユーザ群を Â+群，同様にして C+群との傾向スコアマッチングにより抽出された群を Ã+群とする．なお，傾向スコアマッチングの詳細な手順は付録 2 で述べる．

各評価指標の平均値とウェルチの t 検定の結果を表 4 に示す．まず「興味との適合」に対応する評価指標に着目する．C+群は Ã+群と比較して，継続視聴した動画の割合および動画視聴時間の割合が 0.06 高かった一方で，B+群では統計的に有意な差はなかった．動画視聴開始率は，C+群では有意差はなく，B+群では Â+群より 0.08 低かった．これらの結果から，NMF

ベース手法はユーザの嗜好をよく予測でき，興味に合ったテーマを提示できていたことが示唆された．

次に「予期せぬ発見」に対応する評価指標について論ずる．表 3 より，動画視聴数が両群とも約 2 本増加し，視聴カテゴリ数も有意に増加した．これらの結果から，推薦システムを利用したユーザは利用しないユーザより，予期しないものに出会う可能性が高いと言える．次に，推薦システムが視聴カテゴリの共起に及ぼす影響について分析した．具体的には，非介入群（A 群）で共起率が低いカテゴリ集合について，推薦システムを導入した群（B+群および C+群）での共起率を評価した．これは，5.1 章で述べた通り，A 群で共起率が低いカテゴリ集合が B+群および C+群で多く視聴された場合には，自身にとって予期しなかった動画を視聴したユーザが推薦システムを導入した群で多くなると考えられるためである．

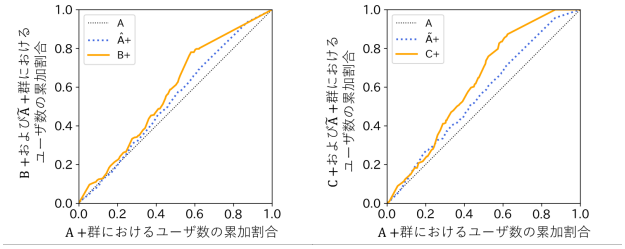
以上のアイデアに基づき，共起カテゴリの分布の比較結果を図 3 に示す．横軸は，A 群において共起率が閾値以下のカテゴリ集合を視聴したユーザの累加割合を表し，縦軸は，対応する A 群のユーザが視聴したカテゴリ集合と同じカテゴリ集合を視聴したユーザの累加割合を表す．例えば横軸の値が 0.2 である場合，B+群の縦軸の値は，A 群のユーザの 20%が視聴したカテゴリ集合を視聴したユーザの割合を表す．従って，評価対象群のユーザが A 群では共起率が低いカテゴリ集合を多く視聴した場合，プロットは左上に現れる．図 3a の Â+ 群および B+を比較すると，B+群の方が共起率が低いカテゴリ集合を視聴するユーザの割合が大きいことが分かった．同様に，図 3a の Ã+ 群および C+群を比較すると，横軸の値が 0.2 付近を除き C+群の方がユーザの割合が大きかった．また，各群の AUC（Area Under the Curve）の値は，Â+ 群が 0.54, B+群が 0.58, Ã+ 群が 0.56, C+群が 0.63 であった．C+群の方が B+群より AUC が大きかった理由は，B+群では人気の動画を推薦するため，視聴されたカテゴリ集合が偏っていたためと考えらえる．一方で，C+群は個人化推薦により，多くのユーザが様々なカテゴリ集合の動画を視聴していたと推測できる．また，複数カテゴリの動画を視聴したユーザの割合は，Â+ 群が 0.69, B+群が 0.80, Ã+ 群が 0.72, C+群が 0.87 であった．以上の結果から，推薦システムはユーザが様々なカテゴリや動画を探索することを促し，予期せぬ発見に貢献することが示唆された．

5.2.4 推薦システムが効果的なユーザの分析

最後に，サイト滞在時間に基づいてユーザを分類し，推薦システムが効果的なユーザの特徴について分析する．サイト滞在時間が長いユーザは，多くの動画を視聴することで網羅的に情報収集しようとしている可能性が高い．一方で，サイト滞在時間が短いユーザは，効率的に自身が興味のある動画だけを視聴しようとしている，または，REX 自体に関心が低いユーザの可能性が高い．そこで，サイト滞在時間に基づいてユーザを分類して評価することで，視聴目的と推薦システムのもたらす効果について分析が可能であると考えられる．具体的には，A/B/C 群においてサイト滞在時間が上位 30%，中間 40%，下位 30%の層にユーザを分類し，5.2.3 章で述べた傾向スコアマッチングを施したユーザ群について評価した．各層の平均サイト滞在時間は図 4a に，ユーザ数は表 5 に示す通りである．

表 4: 傾向スコアマッチングを適用した場合の評価結果．検定は Â+ 群と B+群, Ã+ 群と C+群の比較結果を記載．（$^{**}p < 0.05$; $^{***}p < 0.01$）

	Â+群	B+群	Ã+群	C+群
ユーザ数	291	291	236	236
動画視聴開始率	0.82***	0.74	0.82	0.80
継続視聴した動画の割合	0.83	0.83	0.86	0.92***
動画視聴時間の割合	0.76	0.75	0.77	0.83***
動画視聴数	8.5	10.1**	8.2	10.4***
視聴カテゴリ数	3.0	3.7***	3.0	3.5***
サイト滞在時間（分）	78.6	84.9	72.6	92.5***
イイねクリック率	0.26	0.27	0.27	0.27
アンケート回答率	0.52	0.51	0.57	0.57



(a) Â+ および B+群の比較結果　(b) Ã+ および C+の比較結果

図 3: 共起カテゴリからの乖離度合いの比較結果．プロットが左上にあるほど，非介入群である A 群で一緒に視聴された頻度が低いカテゴリ集合を視聴したユーザの割合が大きいことを意味する．黒点線は A 群同士で比較した場合を表す．

表 5: サイト滞在時間別の各群のユーザ数

層	Â+ 群	B+群	Ã+ 群	C+群
上位 30%	102	124	77	112
中間 40%	121	119	108	97
下位 30%	68	48	51	27

サイト滞在時間の層別の各評価指標の比較結果を図 4 に示す．まず，図 4b~ 図 4d に示した興味との適合に相当する評価指標に着目する．動画視聴開始率は，ほぼ全ての群でサイト滞在時間が短くなるにつれて緩やかに低下し，下位 30%層の B+群のみ極端に低い値となった．継続視聴した動画の割合は，サイト滞在時間が短くなるにつれて低下している．しかしながら，C+群ではその低下傾向が緩やかであり，下位 30%層の C+群は他の群と比較して著しく値が大きい．継続視聴した動画の割合と相関のある動画視聴時間の割合も，下位 30%層の C+群のみ他の群より良い値であった．これらの結果から，ユーザの興味に適合した動画を提示できることは，サイト滞在時間が短いようなユーザに特に効果的であることが分かった．これは，サイト滞在時間が短いユーザは効率的に動画を探索したいというニーズがあり，推薦システムがそのニーズに応えられることを示唆している．

最後に，図 4g および図 4h に示した，推薦結果の精度に着目

する．サイト滞在時間が短いユーザほど，CTR の値が大きくなった一方で，CVR の値は低下し，その低下傾向は B+群の方が顕著であった．このことから，サイト滞在時間が短いユーザほど推薦システムの利用に積極的であり，特にサイト滞在時間下位 30%層には人気に基づく推薦よりも個人化推薦の方が利用率が高いことが分かった．サイト滞在時間上位 30%層で CTR が低かった理由として，自身のペースで網羅的に探索したいという目的で訪問したユーザにとって，推薦システムが必要なかったためだと考えられる．

6 おわりに

6.1 結　　論

本稿では，動画視聴型オンラインカンファレンスにおいて推薦システムがもたらす効果を「興味との適合」と「予期せぬ発見」の観点から実証実験を通して検証した．実証実験では，2022 年に開催された動画視聴型オンラインカンファレンスにおいて，約 4000 人以上を対象に A/B テストを実施した．A/B テストでは，非負値行列因子分解を用いた個人化推薦，人気ランキングに基づく非個人化推薦，推薦を表示しない場合で参加者をランダムに 3 つの群に振り分けた．

実証実験の結果，個人化推薦の結果を提示したユーザの方が，興味に適合した動画を提示でき，web サイトにユーザを長く惹きつけられることが分かった．さらに，アソシエーション分析により，予期せぬ発見のある視聴体験をしたユーザが多いことが示唆された．また，サイト滞在時間が短いユーザほど推薦システムの利用に積極的であったことが分かった．特に個人化推薦の結果を提示したユーザの方が推薦システムの利用率が高く，継続視聴した動画の割合の向上が観察された．

6.2 本実験の限界と今後の課題

本研究には次のような限界があることも付記しておく．

a） 参加者の多様性

本研究で対象とした動画視聴型オンラインカンファレンスは国内で開催されたものであり，様々な文化的背景を持つ参加者を含む国際会議においても同様の結果が得られるかは検証が必要である．また，多くの参加者が特定の業界のグループ会社や関連会社に所属し，働き盛りの男性に偏っていた．カンファレンスで扱う分野や種類，期間，参加者数が異なる場合にも同様の結果が得られるかを検証することは今後の課題である．

b） 推薦システムの非利用者とバイアス除去

本研究では実際に推薦システムを利用したかどうかは制御できず，推薦システムを導入した群では 20% 強の利用率であったため，サンプル数に偏りが生じた．推薦システムを利用しなかった理由として，推薦結果に満足できなかった，あるいは，レイアウトが適切でなく，推薦の提示に気付かなかったことが考えられる．しかしながら，これらの要因を区別することは困難であるため，推薦システムを利用しなかったユーザの分析は本稿の対象外としている．また，推薦システムを利用したユーザに対しても，UI やレイアウトのバイアス [31] やレストルフ効果 [33] を除去することはできず，今後の課題である．

上記のような限界があるものの，本研究は動画視聴型オンラインカンファレンスにおける推薦システムの効果を検証した初めての試みであり，推薦システムの新たな活用先の展望と導入時の知見を得ることができた．本研究では，古典的な推薦手法を適用することで推薦システムの導入が有望であるかを検討したが，動画視聴型オンラインカンファレンスに特化した推薦手法の開発も今後の重要な課題である．

文　　献

[1] D. Salomon, M. F. Feldman, The future of conferences, today: Are virtual conferences a viable supplement to “live” conferences?, EMBO reports 21 (7) (2020) e50883.

[2] T. Bousema, P. Selvaraj, A. A. Djimde, D. Yakar, B. Hagedorn, A. Pratt, D. Barret, K. Whitfield, J. M. Cohen, Reducing the carbon footprint of academic conferences: The example of the American Society of Tropical Medicine and Hygiene, The American Journal of Tropical Medicine and Hygiene 103 (5) (2020) 1758.

[3] S. Sarabipour, A. Khan, S. Seah, A. D. Mwakilili, F. N. Mumoki, P. J. Sáez, B. Schwessinger, H. J. Debat, T. Mestrovic, Evaluating features of scientific conferences: A call for improvements, BioRxiv (2021) 2020–04.

[4] A. Valenti, G. Fortuna, C. Barillari, E. Cannone, V. Boccuni, S. Iavicoli, The future of scientific conferences in the era of the COVID-19 pandemic: Critical analysis and future perspectives, Industrial health (2021).

[5] M.-H. Jang, E.-Y. Choi, How Will Video Conference Fatigue Affect Participants of MICE in the With-COVID-19 Era? Focusing on Video Conference Quality, Social Presence Theory, and Flow, International Journal of Environmental Research and Public Health 19 (8) (2022) 4601.

[6] V. Kalia, A. Srinivasan, L. Wilkins, G. D. Luker, Adapting scientific conferences to the realities imposed by COVID-19 (2020).

[7] P. Resnick, H. R. Varian, Recommender systems, Communications of the ACM 40 (3) (1997) 56–58.

[8] G. Linden, B. Smith, J. York, Amazon. com recommendations: Item-to-item collaborative filtering, IEEE Internet computing 7 (1) (2003) 76–80.

[9] C. A. Gomez-Uribe, N. Hunt, The netflix recommender system: Algorithms, business value, and innovation, ACM Transactions on Management Information Systems (TMIS) 6 (4) (2015) 1–19.

[10] J. Liu, P. Dolan, E. R. Pedersen, Personalized news recommendation based on click behavior, in: Proceedings of the 15th international conference on Intelligent user interfaces, 2010, pp. 31–40.

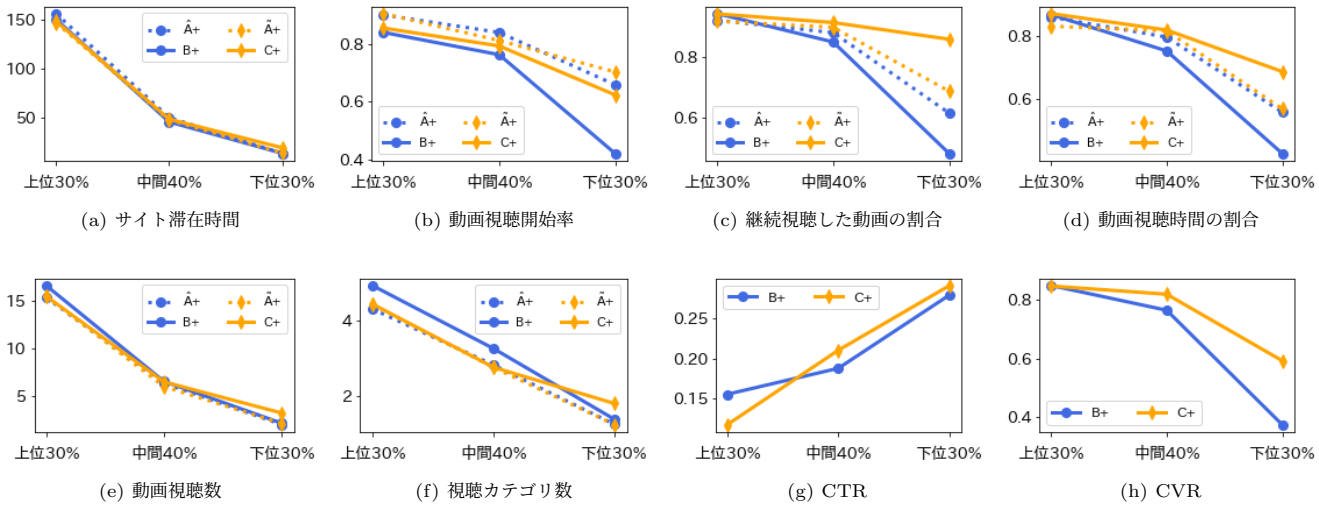
[11] J. Lu, D. Wu, M. Mao, W. Wang, G. Zhang, Recommender system application developments: A survey, Decision Support Systems 74 (2015) 12–32.

[12] C. Koo, N. Chung, J. Ham, Assessing the user resistance to recommender systems in exhibition, Sustainability 9 (11) (2017) 2041.

[13] 大社綾乃, 大滝啓介, 石井良尚, 中島健介, 川崎滉平, 小出智士, オンラインカンファレンスにおける情報推薦導入の検討, in: 第 14 回データ工学と情報マネジメントに関するフォーラム, F34-2, 2022.

[14] R. Kohavi, D. Tang, Y. Xu, Trustworthy online controlled experiments: A practical guide to A/B testing, Cambridge University Press, 2020.

[15] X. Liu, R. Seevers, Z. Gu, X. Yang, Smart MICE: Definitions, foundations and development, in: 2020 7th International Conference on Information Science and Control En-



(a) サイト滞在時間　(b) 動画視聴開始率　(c) 継続視聴した動画の割合　(d) 動画視聴時間の割合

(e) 動画視聴数　(f) 視聴カテゴリ数　(g) CTR　(h) CVR

図 4: サイト滞在時間の層別の各評価指標の比較結果.

gineering (ICISCE), IEEE, 2020, pp. 1296–1300.

[16] A.-P. Correia, C. Liu, F. Xu, Evaluating videoconferencing systems for the quality of the educational experience, Distance Education 41 (4) (2020) 429–452.

[17] S. B. Alzahrani, A. A. Alrusayes, M. S. Aldossary, Impact of COVID-19 pandemic on dental education, research, and students, Int J Health Sci Res 10 (6) (2020) 207–12.

[18] K. A. Karl, J. V. Peluchette, N. Aghakhani, Virtual work meetings during the COVID-19 pandemic: The good, bad, and ugly, Small Group Research 53 (3) (2022) 343–365.

[19] L. Fosslien, M. W. Duffy, How to combat zoom fatigue, Harvard Business Review 29 (2020) 1–6.

[20] A. Kuzminykh, S. Rintel, Low engagement as a deliberate practice of remote participants in video meetings, in: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–9.

[21] A. Oruc, Tools for organizing an effective virtual academic conference, Serials Review 47 (3-4) (2021) 231–242.

[22] J. Foramitti, S. Drews, F. Klein, T. Konc, The virtues of virtual conferences, Journal of Cleaner Production 294 (2021) 126287.

[23] O. Reshef, I. Aharonovich, A. M. Armani, S. Gigan, R. Grange, M. A. Kats, R. Sapienza, How to organize an online conference, Nature Reviews Materials 5 (4) (2020) 253–256.

[24] Ö. Yozcu, H. Kurgun, D. Bağıran, Factos that influence attendance, satisfaction and loyalty for virtual events, Advances in Hospitality and Tourism Research (AHTR) (06 2022). doi:10.30519/ahtr.1068444.

[25] P. Lops, M. d. Gemmis, G. Semeraro, Content-based recommender systems: State of the art and trends, Recommender systems handbook (2011) 73–105.

[26] J. B. Schafer, D. Frankowski, J. Herlocker, S. Sen, Collaborative filtering recommender systems, in: The adaptive web, Springer, 2007, pp. 291–324.

[27] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (8) (2009) 30–37.

[28] R. Agrawal, T. Imieliński, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of the 1993 ACM SIGMOD international conference on Management of data, 1993, pp. 207–216.

[29] S. Vargas, P. Castells, Rank and relevance in novelty and diversity metrics for recommender systems, in: Proceedings of the fifth ACM conference on Recommender systems, 2011, pp. 109–116.

[30] M. Ge, C. Delgado-Battenfeld, D. Jannach, Beyond accuracy: evaluating recommender systems by coverage and serendipity, in: Proceedings of the fourth ACM conference on Recommender systems, 2010, pp. 257–260.

[31] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, X. He, Bias and debias in recommender system: A survey and future directions, arXiv preprint arXiv:2010.03240 (2020).

[32] D. B. Rubin, N. Thomas, Matching using estimated propensity scores: Relating theory to practice, Biometrics (1996) 249–264.

[33] R. R. Hunt, The subtlety of distinctiveness: What von Restorff really did, Psychonomic Bulletin & Review 2 (1) (1995) 105–112.

付　　録

1　イベントログから動画視聴時間の算出

ユーザ毎にイベントログを時系列で並び替え，ページの訪問および更新のログが連続している場合は，最初のログのみを残し，動画視聴開始ボタンのクリック等の他のログが連続している場合は，最後のみを残した．時系列順に並び替えた各ユーザのログから，動画視聴開始ボタンのクリックから次のイベントまでの時間を動画視聴時間とした．なお，動画視聴時間が動画の実時間より長い場合は，動画の実時間を動画視聴時間とした．

2　傾向スコアマッチングの手順

A 群と B+群で選択バイアスを緩和するための傾向スコアマッチング [32] の具体的な手順は以下の通りである．まずはじめに，B 群のログデータを用いて，ユーザが推薦システムを利用する確率を予測するロジスティック回帰モデルを構築する．この確率を傾向スコアと呼ぶ．共変数として，イイねクリック率とアンケート回答率を用いた．次に，構築したモデルを A 群のデータに適用し，各ユーザの傾向スコアを計算する．最後に，A 群と B ＋群のユーザを傾向スコアの最近傍マッチングさせる．具体的には，傾向スコアが 0.5%以内で一致するユーザを両群からランダムに選択し，残りのユーザは両群から削除する．C+群も同様の手順で実施した．