

動画閲覧時の動作特性によるタスクタイプの分類と動画推薦 Classification of Video Viewing Task Types and Recommendation of Videos

井出 達郎
Tatsuro Ide

法政大学情報科学部コンピュータ科学科
E-mail: tatsuro.ide.2f@stu.hosei.ac.jp

Abstract

YouTube is one of the largest and most sophisticated recommender systems and a useful source of information for users. In video search on YouTube, even the same user has different purposes in mind depending on the user's state. However, videos are recommended based on the relevance of videos and the user's view history, regardless of the user's state. This paper proposes a classification of video viewing task types based on the user's behavioral characteristics. By classifying the user's purpose as a task type, it enables higher-order recommendation that fits the task type. Behavioral characteristics are momentary characteristics of the user that appear from actions such as screen scrolling and video previewing. The system implicitly records the user's actions, classifies the task type based on these parameters, and recommends the related video list on a mobile application that imitates YouTube. Experiments were conducted to evaluate classification accuracy of task types and recommendation. We conducted a pre-questionnaire on YouTube usage and a post-questionnaire on video recommendation for each participant. This paper presents the points to be improved and discussions on future research.

1. はじめに

YouTube は、現存する最大規模で最も洗練された推薦システムの 1 つであり、ユーザにとって有用な情報のリソースである。YouTube 上での動画探索において、同じユーザであっても、状態によって心に抱いている目標は異なる。しかし、現状はユーザのその時々状態に関係なく、動画の関連度や過去の視聴履歴から動画が推薦されていると推測される。例えば、学習目的でコンテンツを探しているときであっても、登録しているチャンネルの動画や集中力を欠くようなコンテンツが推奨されてしまうことがある。また、探索的に幅広い動画を求めているときであっても、普段見ている動画群に戻ってしまい、新しい発見に行き着かないことがある。このように雑多に関連動画を表示することで、ユーザの当時の思惑とは異なる行動につながったり、閉塞的な検索空間に閉じこもってしまったりすることは多い。

本研究では動画閲覧時の動作特性によるタスクタイプの分類とタスクタイプ毎の動画推薦を提案する。動作から読み取れる動的なユーザの目的をタスクタイプとして分類することで、その目的に沿った高次の推薦をすることが狙いである。動作特性とは、画面のスクロールや、動画プレビューなどの細かなユーザの動作から現れるそのユーザの瞬間的な特徴である。YouTube を模したモバイルアプリケーション上で暗黙に記録したユーザの動作によるパラメータを基にタスクタイプの判定をして、関連動画表示で動画推薦を行う。タスクタイプの分類精度と推薦の評価をするために実験を行う。各参加者の YouTube 利用についての事前アンケートと、動画推薦についての事後アンケートを行った。最後に、改善点と今後の課題についての議論を述べる。

2. 関連研究

Athukorala ら [1] は、Google Scholar のようなインタフェースを有する論文検索エンジンにおいて、動作特性によるタスクタイプの分類を行った。この論文では、調査か探索かの 2 つのタスクタイプの分類が行われており、動作特性として、クエリの長さ、読み取り時間、累積クリック数を利用している。これらを基にしたパラメータによってタスクタイプを判定する分類器を構築した。Google は YouTube の推薦アルゴリズムで、暗黙的な特徴量を Deep Learning の訓練に利用した [2]。暗黙的な特徴量として、過去の視聴履歴や視聴地域、性別などが挙げられた。

3. 準備

本研究では決定木の生成に、Weka [3] の有する機械学習アルゴリズムの中で分類学習のアルゴリズムである C4.5 [4] を利用する。データセットとして著者自身にタスクを課して得たデータを利用し、入力変数としてクエリの長さ、読み取り時間、スクロール深度、取りうる出力としてタスクタイプを記述する。C4.5 は情報理論に基づき、決定木の葉の識別力の定義として平均情報量を用いる。データセットの集合 C において、取りうる出力は集合 D に属するものとして、 $x \in D$ が起こる割合を $p_x(C)$ と表す。データセットの集合 C に対する平均情報量 $M(C)$ は以下になる。

$$M(C) = - \sum_{x \in D} p_x(C) \log p_x(C)$$

次に C をある入力変数において分割することで、平均情報量 M' を求める。先に求めた平均情報量 $M(C)$ との差を情報ゲインという。これは入力変数、つまり動作特性によるパラメータによって変化した情報量を表し、これが最も大きいものを識別力の高い条件としてノードに選択し、これを他ノードに対しても再帰的に行うことによって決定木を生成する。

4. 提案手法

本研究では、ユーザの動作特性から読み取れる動的なユーザの目的をタスクタイプとして分類し、推薦に応用する。タスクタイプは Athukorala らの論文 [1]をもとに定義する。この論文では、論文検索において調査と探査の2つのタスクタイプを定義した。本研究では動画探索時について定義するため、これに周回を加えて以下の3つを定める。

- 調査：あらかじめ探す動画が決まっているタスク。ユーザは特定の動画を目標として検索する。
- 探査：探す動画が決まっていないタスク。ユーザは興味や関心から、幅広いコンテンツを検索する。
- 周回：同じようなコンテンツを繰り返し見るタスク。

動作特性とは、瞬間的なユーザの特徴を表す行動のことである。本研究では、以下の3つの動作特性を記録し、パラメータとして利用した。

- クエリの長さ：最初の検索セッションでのクエリに入力した単語数。スペース区切りでカウントする。
 - 画面スクロール深度：動画リストのビューを上下にスクロールした深さ。
 - 読み取り時間：最初の動画を見始めるまでの時間。
- 提案手法は主に UI, 分類器, 推薦の3段階からなる。
1. UI は動画閲覧時のユーザの動作を記録するものであり、YouTube クライアントアプリケーションとして実現する。これは YouTube mobile のものほとんど変わらないが、暗黙に動作の記録を行う。
 2. これらの動作からパラメータを抽出し、分類器に渡す。分類器はパラメータを受け取ってタスクタイプを判定する。
 3. タスクタイプによって動画検索にフィルタを掛ける形で推薦を行う。

4.1. 分類器のパラメータ設定

本研究では、動作特性のパラメータ選定に Weka を用いる。データセットとして本アプリケーション上でタスクを著者自身に課すことで得たデータを利用する。自身に課したタスクは、調査タスク6個、探査タスク6個、周回タスク8個の合計20タスクである。調査タスクでは、思いついた特定の動画を探索した。探査タスクでは、カテゴリを決めて探索した。周回タスクでは、日常的に確認している動画群を探索した。

このデータセットを入力データとして、機械学習アルゴリズムには C4.5 [4]に基づいた決定木を生成する J48 を選択した。このアルゴリズムを採用した理由は、決定木

による優れた可視性から分類失敗の原因を理解しやすいためである。学習データのために交差検証を用いて、生成された決定木は図1である。この決定木は自身で行った20タスクを70%の精度で分類する。子を持つ葉ノードはそれぞれクエリの長さ(query_length)、スクロール深度(scroll_depth)を表す。子を持たない葉ノードは各タスクを表し、調査(lookup)、探査(exploration)、周回(repeat)の3つある。枝の比較演算は各データの親ノードのパラメータによって分岐する。

この決定木の生成で最初の条件としてクエリの長さを選ばれているのは、その情報ゲインがスクロール深度や読み取り時間のものと比べて大きかったためである。情報ゲインは以下のように計算される。データが調査、探査、周回の3つに分かれ、各クラスのデータ数が6個、6個、8個であるので、平均情報量は以下になる。

$$M(C) = \frac{6}{20} \log_3 \frac{20}{6} + \frac{6}{20} \log_3 \frac{20}{6} + \frac{8}{20} \log_3 \frac{20}{8} \approx 0.991$$

さらに入力変数であるクエリの長さについて、0より大きいかという条件で分割して平均情報量を求めると、 $M' = 0.582$ となり、情報ゲインは0.409となる。

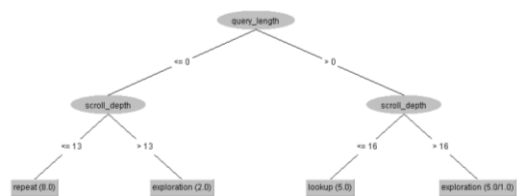


図1 著者自身の操作記録に基づいたタスクタイプの決定木

4.2. 推薦

分類器によって判定されたタスクタイプによって、検索パラメータを変更する。タスクタイプを探査と判定すると、関連動画表示に興味のあるカテゴリの動画を追加する。タスクタイプを調査と判定すると、関連動画に登録チャンネルの新作動画やライブストリーミングを特別に表示せず、動画の関連度の高い順に表示する。タスクタイプを周回と判定すると、動画の関連度と関係なしに、関心の高い登録チャンネルの新作動画を関連動画表示に追加する。

5. 実装

本研究では、アプリケーションの作成に Android Studio を使用し、端末として Google Pixel 4a を用いた。YouTube 動画データは YouTube Data API を用いて取得し、Android Player APIによって再生する。このアプリケーションは、YouTube 上の動画を閲覧する際のユーザの動作特性を記録し、タスクタイプを分類し、推薦に適用する。動作として、クエリの長さ、読み取り時間、画面スクロール深度を記録する。UI は YouTube mobile のものを参考にした。有効な下部のタブはホーム、探索、登録チャンネルで、事前アンケートによって各被験者に対応した動画リストが検索される。画面上部の虫眼鏡アイコンを押すことで、

検索アクションが始まる。各タブや検索結果の動画リストから動画をタップすると動画が再生される。動画再生と関連動画表示のスクリーンショットを図2に示す。



図2 実装したアプリケーション

6. 実験

この実験の目的は、動作特性によるタスクタイプの動的な分類と推薦の評価をすることである。本システムは3つのユーザインタラクションのデータ(クエリの長さ、スクロール深度、読み取り時間)を記録する分類器を使う。また、分類器はこれらの入力データをパラメータとして扱い、ユーザのタスクが調査か探索か周回か予測する。ここでは、この分類器が有効なシステムのことをフルシステムと呼ぶ。この分類器を有効にしないシステムのことをベースラインシステムと呼び、比較として使う。

参加者は4名で、平均年齢は22.3歳、日常的に何らかの用途でYouTubeを利用している。YouTubeの各タブを疑似的に再現するために、事前アンケートを行った。

6.1. 設計

被験者1人당りに調査タスクを1タスク、探索タスク、周回タスクをそれぞれのフルシステム、ベースラインシステムで行うため4タスク、合計5タスクをこなしてもらう。また、順序効果を避けるためにタスクの順番はバランスをとる。各タスクについてシチュエーションを設定する。

調査タスクは実験の2時間前に見てもらった動画を本アプリケーションで探してもらう。探索タスクは、事前アンケートで答えてもらった興味はあるが知識はあまりない分野をいくつか答えてもらったうえで、適切なものを選択し、自由に調べてもらう。表1は探索タスクで答えてもらったカテゴリと参加者の対応を示す。周回タスクは、日常的に見ている動画群を事前に答えてもらい、いつも通りに利用してもらう。探索タスク2つと周回タスク2つがそれぞれ終わったタイミングで前後のシステムの関連動画表示について質問をする。これもフルシ

テムとベースラインシステムで順番を実験者によって変える。

表1 事前アンケートで選ばれた興味のあるカテゴリから実験用に選んだ検索トピック

参加者	検索トピック
1	映画とアニメ
2	科学と技術
3	音楽
4	映画とアニメ

6.2. 手続き

実験の手続きは事前アンケート、調査タスクのための動画視聴、本実験の3段階に分かれる。事前アンケートでは、疑似的に個別のYouTubeページを再現するためにいくつかの質問を用意した。ホームタブと登録チャンネルタブのために、よく見る動画カテゴリと登録チャンネルのリストを答えてもらった。さらに、探索タスクで探索してもらうカテゴリのために、興味はあるが知識が限られていると認識しているカテゴリも複数答えてもらい、その中でAPIから有効な結果が得られるものを選択して、推薦に反映できるようにした。

調査タスクのシチュエーションを被験者毎に再現するために、実験の2時間前に個別で特定の動画を視聴してもらった。コンテンツを詳細に記録することを防ぐために、後に行う調査タスクのことは伝えなかった。この動画は5分程度のもので、見終わったらそれぞれの作業に戻ってもらい、2時間後に実験を行うことを伝えた。時間を空けることで、タイトルや投稿者などの詳細な情報を忘れるようにした。

2時間たったあとで本実験の前に、アプリケーションの機能とタスクについて説明した。各参加者は調査タスク1つとそれぞれのシステムに対して探索タスクと周回タスクの組み合わせで4つの合計5タスク行ってもらい、調査タスクは、最大15分で動画が見つかるまで探索してもらい、参加者は動画を見つけた時点でタスクを終了できる。探索タスクと周回タスクは20分間ずつ探索に費やしてもらう。本実験はそれぞれ90分ほどかかった。

6.3. ユーザの評価

タスク全体を通して、関連動画表示の差異に気付かなかったと答える参加者が多かった。関連動画表示に目的と全く異なる動画表示が出ていることに気付いたという声があった(参加者3)。これは本来のタスクと異なるタスクで判定されたために、関連動画表示に目的と全く関係のない動画が含まれていたためである。他にも、同じ動画が繰り返し表示されたことと答えた参加者もいた(参加者4)。これも推薦として関連動画表示に追加された動画に違和感を覚えたための回答である。

6.4. 分類器の精度

参加者1人당りに5タスクで、4人に取り組んでもらったため、合計20タスクについて精度を計算する。20タスクのうち正しく判定で来たのは10タスクで精度は50%

であった。このうち、調査タスクは4タスクで、2タスク正しく判定できた(50%)。探査タスクは8タスク中8タスクが正しく判定されたため100%、周回タスクは全体で一度も判定されなかったため0%となった。また、周回タスクの8タスクのうち7タスクが探査タスクと誤って判定された。

正しく分類されなかったタスクについて分析すると、著者自身の操作で集めたデータと比べて、参加者のデータでは、全体的にスクロール深度が大きかったこと、周回タスクでのクエリ検索もあったことが精度の低さにつながっていた。操作に慣れていないことや、機能の確認を最初に割り当てられたタスクで行うことで、最初の検索セッションでのスクロールがシステムに多くカウントされていた。また、自身の操作で周回タスクを行った際には、登録チャンネルタブから見る動画を決めたため、検索クエリがすべて0になり、結果として生成された分類器ではクエリの長さが1以上で周回タスクと判定されるケースがなくなっていた。

6.5. ユーザデータを基にした決定木

今回の実験で得た参加者の動作特性パラメータと与えられたタスクを入力データとして、Wekaで新たに生成した決定木を図3に示す。交差検証によるものであるため単純に比較できないが、本システムの分類器に採用されている著者自身のデータのみで生成した決定木での分類精度が50%であるのに対して、こちらの決定木は70%である。分類器に採用した決定木では、クエリの長さが0よりも大きい、すなわち検索をしたときに判定されるタスクは探査と調査しかないため、検索して周回タスクを行うと必ず誤ってタスクタイプの判定が行われてしまう。一方でユーザデータを基にした決定木ではクエリの長さが0でないときでも周回タスクと判定されるケースが存在する。また、スクロール深度は自身の操作によるデータよりもユーザデータの方が全体的に大きかったため、スクロール深度による閾値に大きい差がある。

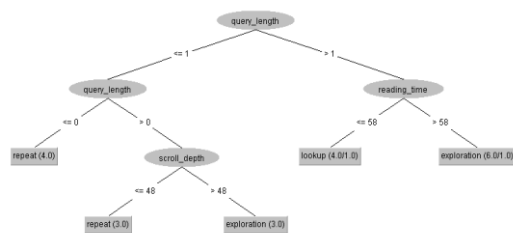


図3 参加者の動作特性によって生成された決定木

7. 議論

実験により、調査タスクと探査タスクについては、一定程度の精度で分類できることが分かった。一方で周回タスクに分類されるケースはなかったこと、全体的に探査タスクと判定されることが多かったことから、偏った結果になったといえる。まず考えられる原因は、決定木生成のために用意したデータが著者自身にタスクを課し

たものみであったことである。探査タスクと判定されるケースはクエリの長さが0以下でスクロール深度が13より大きいとき、クエリの長さが0より大きくスクロール深度が16より大きいときである。つまりクエリの長さに寄らずある一定以上スクロールすると探査タスクと判定される。自身にタスクを課して動作を記録した際には、全体的にスクロール深度が小さく、これに対して実験の参加者のスクロール深度は全体的に大きかった。この差は、自身の各タスクに対する先入観や実験の参加者数が少ないことによって生まれたように考えられる。この問題に対しては、システムに組み込む決定木を生成するデータを得るために予備実験をする、本実験数を増やす等、データの偏りに配慮する必要がある。

YouTubeの機能を再現するにあたって、APIで実装できる部分と疑似的な実装が必要な部分があった。探索タブの急上昇動画のリストやクエリ検索などは現在のAPIの機能で実装できるが、ホームタブに表示される各ユーザーのおすすめ動画のリストは以前のAPIバージョンでは取得できたが、現在は撤廃された機能である。他にも登録チャンネルのリストを手に入れることができても、投稿日時順に動画リストを取得することができない。本研究では、これらの利用できない機能を代替するために事前アンケートを行い、普段利用しているYouTubeを疑似的に再現した。

8. おわりに

動的にユーザのタスクタイプを分類するために動作特性によるパラメータを記録し、決定木によって分析した。暗黙にクエリの長さ、スクロール深度、読み取り時間を記録し、決定木によってタスクタイプを判定し、関連動画表示に反映するアプリケーションを作成することができた。関連動画表示に推薦をすることで、ユーザに何も促さないと動画リストの差異に気付けないことが問題であった。一方で関連動画表示の差の認知を促すと不自然に関連動画表示を確認してしまうため、スクロール深度が非自然に深くなってしまう可能性がある。Athukoralaら[1]のようにスクロール深度を採用せずにブックマークなどで暗黙に検索結果リストを評価する等、評価手法の改良が今後の課題である。

文献

- [1] K. Athukorala, A. Medlar, A. Oulasvirta, G. Jacucci and D. Glowacka, "Beyond Relevance: Adapting Exploration/Exploitation in Information Retrieval," *Proc. ACM IUI*, pp. 359-369, 2016.
- [2] P. Covington, J. Adams and E. Sargin, "Deep Neural Networks for YouTube Recommendations," *Proc. RecSys*, pp. 191-198, 2016.
- [3] I. H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, San Francisco: Morgan Kaufmann, 2005.
- [4] J. Quinlan, *C4.5: Programs for Machine Learning*, Los Altos, California: Morgan Kaufmann, 1993.