

チュートリアル動画のための理解容易性因子の抽出

樽見 彰仁[†] 北山 大輔[†]

[†] 工学院大学情報学部 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: tj115071@ns.kogakuin.ac.jp, tikitayama@cc.kogakuin.ac.jp

あらまし Youtube などの動画共有サイトにはソフトウェアの学習を支援するためのチュートリアル動画が数多く存在する。チュートリアル動画はソフトウェアとその操作目的ごとに存在している。チュートリアル動画は、視聴してからではないと理解が容易であるか不明であり、学習に適している動画を発見することは容易ではない。そこで、本研究では理解が容易なチュートリアル動画を定義するため、重要な理解容易性因子を操作の解説量と動画内で使用されている編集技法から抽出する。具体的には、クラウドソーシングを用いて、データセットとしたチュートリアル動画に対して、わかりやすさをアノテーションする。そのアノテーションを目的変数とし、理解容易性因子の候補を説明変数とする回帰分析を行うことで、重要な理解容易性因子を決定する。

キーワード チュートリアル動画、動画共有サイト、ソフトウェア学習支援、回帰分析、クラウドソーシング

1 はじめに

ユーザは操作が複雑なソフトウェアを利用する際、操作方法や技法を学ぶために、YouTube¹などのインターネット上の動画共有サイトにアップロードされたチュートリアル動画を参考にすることがある。チュートリアル動画は、各ソフトウェア、目的ごとに存在し、内容は実際にソフトウェアを操作したものをキャプチャし、それを字幕や音声などで解説したものとなっている。それらのチュートリアル動画は、ひとつの操作ごとに詳細な解説をしているものから、ソフトウェア上で操作しているものをキャプチャしているのみで解説がないものなど幅広く存在している。動画内の編集技法（トランジション [1]、テロップの表示方法 [2]）や動画の構成 [3] [4] などが視聴者に与える影響は大きく、チュートリアル動画においても視聴や理解を容易にするために、マウスカーソルの強調、ショートカットキーの表示、メニュー選択時に拡大、同じ操作の時間が長い場合に早送りやスキップなどの様々な視覚的な工夫がなされている。

本研究では多数存在するチュートリアル動画の中から、操作の音声による解説量が適切であり、視聴しやすい工夫がなされている動画を定義するために、重要な理解容易性因子の抽出を行う。このことによりソフトウェアの学習を支援することを目的としたチュートリアル動画検索を実現することができる。本稿では対象データとして YouTube にアップロードされた Photoshop のチュートリアル動画を対象とした。

操作の解説量はソフトウェアの操作に関する単語がどの程度解説されているかを動画字幕から算出する。動画内での工夫については人手でアノテーションを付与し、動画の理解容易性については、クラウドソーシング上でタスクを作成し、ワーカーに2つの動画を提示して、どちらの動画が学習に向いているかを比較させる。これらの理解容易性因子の候補を説明変数とする回帰分析を行い、重要な理解容易性因子を決定する。

本稿の構成を以下に述べる。2 節では、利用する理解容易性因子とその抽出方法について述べる。3 節では、クラウドソーシングを用いて、データセットとしたチュートリアル動画に対して、わかりやすさをアノテーションする。そのアノテーションを目的変数とし、抽出した理解容易性因子を説明変数とした回帰分析を行い、結果と考察について述べる。4 節では、動画の編集や構成、ソフトウェア学習支援に関する研究の紹介を行う。5 節では、まとめと今後の課題について述べる。

2 理解容易性因子

2.1 視覚的容易性

理解容易性因子のうち視覚的な容易性として、以下の4つの動画内の工夫を用いる。

- メニュークリック時のズーム

図1のように、クリックしたメニュー周辺に対して、ズームがされているか否かを用いる。

- マウスカーソルの強調

図2のように、マウスカーソルの位置を着色等で強調されているか否かを用いる。

- ショートカットキーの表示

図3のように、ショートカットキーを利用した際、字幕等で表示されているか否かを用いる。

- 単純操作を繰り返す際の早送り、スキップの適切な利用対象を選択するといった、単純操作を繰り返す際、早送りやスキップ等の工夫がなされているか否かを用いる。

2.2 意味的容易性

意味的な容易性として解説量を用いる。動画内においてソフトウェアで使われている単語がどれだけ解説されているかを解説量として算出する。ソフトウェアで使用されている単語の辞書として Photoshop のメニューやツール名を利用した。

以下の手順で解説量の算出を行う。

¹ : <https://www.youtube.com/>

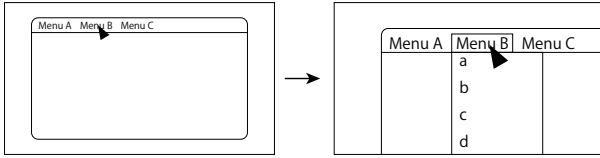


図 1 メニュークリック時のズーム

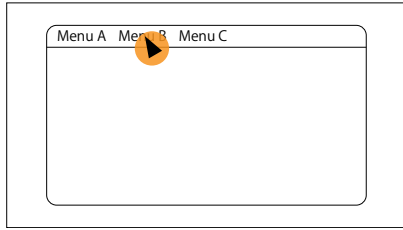


図 2 マウスカソールの強調

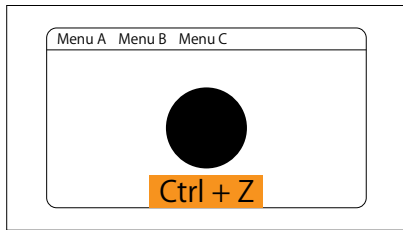


図 3 ショートカットキーの表示

1. YouTube からチュートリアル動画の字幕を取得する。
2. すべての単語のスコアを 1 とする。
3. 字幕内で出現し、かつ辞書中の単語ならば、スコアに χ^2 値を加算する。
4. 単語スコアを合計した値を動画の解説量とする。

2.2.1 前 処 理

YouTube から字幕を取得し、字幕が音声からの自動生成である場合、2 つの単語が 1 つに結合されてしまう場合がある。この問題を解決するために、結合された単語を先頭から区切り、前半部分と後半部分の両方が言語処理ライブラリの辞書²中に含まれたとき単語を分割する。

“selectedlayer” という単語の分割の例を図 4 に示す。初期状態ではポインタが 0 の位置にある。空文字と “selectedlayer” という文字は辞書に含まれていないのでポインタを進める。“select” と “edlayer” に分割したとき、“select” は辞書に含まれているが、“edlayer” は辞書に含まれていないため、ポインタをさらに進める。“selected” と “layer” 分割にしたとき、両方の文字列が辞書に含まれているため分割を完了する。

2.2.2 単語のスコア付け

まず、すべての単語のスコアを 1 で初期化する。一般的によく使用される動作については、解説が少ない傾向があるため操作によって重み付けをする必要がある。そのため本研究では、対象の単語に対して重みとして、 χ^2 値 (式 1) を加算する。 $x_{i,j}$ を対象ソフトウェアのある操作目的のチュートリアル動画 j における単語 i の出現回数とし、 $m_{i,j}$ を動画 j における単語 i の期待値 (式 2) とする。期待値の第 1 項は対象ソフトウェアに関

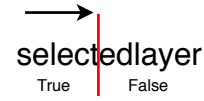
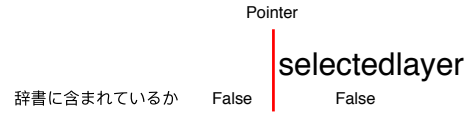


図 4 単語分割

表 1 χ^2 値 結果

単語 \ 目的	crop	skin	glitch
crop	4914.0	-	-
retouch	33.9	342.0	-
randomize	-	-	389.9
value	32.4	108.9	58.5

するチュートリアル動画集合での単語 i の出現確率を示し、第 2 項は動画 j の単語数を示している。ここで、 V は、対象ソフトウェアに関する全てのチュートリアル動画の集合であり、 W はその中に出現するすべての単語集合である。

$$\chi_{i,j}^2 = \frac{(x_{i,j} - m_{i,j})^2}{m_{i,j}} \quad (1)$$

$$m_{i,j} = \frac{\sum_{v \in V} x_{i,v}}{\sum_{w \in W} \sum_{v \in V} x_{w,v}} \times \sum_{w \in W} x_{w,j} \quad (2)$$

今回、YouTube にアップロードされている動画を対象とし、Photoshop のチュートリアル動画の全体集合として “Photoshop Tutorial” で検索した関連度順上位 500 件のうち、字幕がある、または字幕生成が可能なもの 310 件の動画を利用した。また同じ目的に対する Photoshop のチュートリアル動画の集合として “Photoshop Tutorial [目的を表すキーワード]” で検索した関連度順上位 20 件のうち、字幕または字幕生成が可能なものを利用した。目的を表すキーワードとして crop, skin, glitch を用いた。利用可能な動画は crop, skin, glitch それぞれ 17 件、12 件、11 件であった。3 つの目的それぞれの χ^2 値を算出した結果の一部を表 1 に示す。“-” はその単語が出現していないことを示している。

このように各目的に対して、適切に重み付けがされていることが確認できる。

2.2.3 解説量スコアの例

目的 crop での実際の解説量スコアの例を表 2 に示す。目的 crop での特徴的な単語である “crop” という単語が含まれる文がよりスコアが高いことが確認できる。

2 : https://spacy.io/models/en#section-en_vectors_web_lg

表 2 解説量スコア例

文	辞書中の単語スコア	その他の単語スコア	解説量
you get the crop tool by clicking on this fifth tool down or you can press C key	4914 (crop)	17	4931
i'm gonna do as well is click on my background and press ok because we want it as an on lock layer	1637 (layer)	21	1658

表 3 質問内容

質問内容 (日本語)	回答
メニューをクリックしたときズームされているか	yes / no
マウスカーソルが強調されているか	yes / no
ショートカットキーを使ったとき表示されているか	yes / no
カット, 早送り, スキップは適切に使用されているか	yes / no

3 動画内の工夫, 理解の容易さについての調査

3.1 調査方法

2つの目的 (crop, glitch) について調査を行う。動画内の工夫については、正確なデータを得るために、各目的 20 件合計 60 件の動画について 12 人の作業者が各動画につき 1 人で、アノテーションを付与した。アノテーション付与のための質問項目を表 3 に示す。また理解の容易性についての調査のために、Mechanical Turk³上でタスクを作成した。内容は 2 つの動画を視聴してもらい、動画の工夫の有無と理解の容易さとして、どちらの動画のほうが学習に向いているかを回答してもらう。各目的ごとに総当たりで、組み合わせ 1 目的あたり 190 組み合わせ⁴を評価した。1 組み合わせに対して 3 人以上有効な回答が集まるよう募集を行った。調査を行うにあたって、ワーカーが Photoshop について知識を持っているか確認する必要があるため、クイズ (図 5) を作成し、10 点中 7 点以上正解である場合のみタスクを行った。回答は、計 655 人のワーカーが行い、同一人物が回答した最大の組み合わせ数は 46 個であった。

調査結果から理解容易性因子の抽出を行う。解説量、動画内の工夫を説明変数、動画の理解の容易さを目的変数とし、重回帰分析を行う。

重回帰分析を行うにあたって説明変数同士の相関係数を求め、相関関係を明らかにする。表 4 に目的 crop, 表 5 に目的 glitch, 表 6 に 2 つの目的の結果を合わせたものの説明変数同士の相関係数を示す。いずれの項目も強い相関がないことが確認できた。

3.2 因子と理解容易性分析

各目的について重回帰分析を行う。各変数、平均が 0、分散が 1 になるよう正規化を行い、目的変数である理解の容易さについては、その動画を含む全組み合わせにおいて、その動画がわかりやすい動画であると答えた割合を利用した。結果を表 7 に示す。

目的 crop では、切片は-5.458, 決定係数は 0.259 であった。

*必須

Worker ID *

回答を入力

What does this icon show in Photoshop? *

1 ポイント



☐ Lasso
☐ Stamp
☐ Eraser
☐ Healing Brush

What does this icon show in Photoshop? *

1 ポイント



☐ Move
☐ Brush
☐ Color Picker
☐ Quick Selection

"Levels" in Photoshop are: *

2 ポイント

☐ Filters that have been applied to the image
☐ Color and brightness correction
☐ Previous versions of an image
☐ Images stacked on top of each other

What tool allows you to copy one area of the layer to another area of the layer? *

2 ポイント

☐ Magic Wand
☐ Clone Stamp
☐ Lasso
☐ Quick Selection

What keys are to undo the last change? *

2 ポイント

☐ CTRL + S
☐ CTRL + D
☐ CTRL + Z
☐ CTRL + U

送信

Google フォームでパスワードを保護しなくても大丈夫。

図 5 Photoshop Quiz

解説量が最も良い影響を及ぼし、解説量に続いてメニューのズーム、スキップの有効活用の順に良い影響を及ぼすことが確認できた。マウスカーソルの強調については、影響が少なく、ショートカットキーの表示については悪い影響を及ぼすことが確認できた。

目的 glitch では、切片は-2.391, 決定係数は 0.566 であった。解説量が最も良い影響を及ぼし、解説量に続いてメニューのズーム、ショートカットキーの表示、マウスカーソルの強調の順に良い影響を及ぼすことが確認できた。スキップの有効利用については、影響が少ないことがわかった。

目的 crop と glitch を合わせた結果では、切片は 1.587, 決定係数は 0.329 であった。解説量が最も良い影響を及ぼし、解説量に続いてメニューのズームが良い影響を及ぼすことが確認できた。マウスカーソルの強調、スキップの有効活用については影響が少ないことが確認できた。ショートカットキーの表示は悪い影響を及ぼすことが確認できた。

以上の結果から、解説量が最も良い影響を及ぼすことがわかった。ショートカットキーの表示については、画面上ではなく、音声で解説している動画が存在し、画面上の表示の有無のみでは正しい結果を得られなかったと考えられる。スキップの有効活用について、目的間での大きな差が確認された原因として、目的 crop では単純操作である選択作業が多く存在し、目的 glitch では、単純操作を用いた作業が少ないことが原因であると考えられる。

3 : <https://www.mturk.com/>

4 : 先に提示した動画がわかりやすいと回答された割合は 51.4%であり、提示順による影響はないものと考えている。

表 4 各説明変数の相関係数 (crop)

	解説量	メニューのズーム	マウスカーソルの強調	ショートカットキーの表示	スキップの有効利用
解説量	1.000	-0.007	-0.095	-0.173	-0.215
メニューのズーム	-0.007	1.000	-0.236	0.356	-0.272
マウスカーソルの表示	-0.095	-0.236	1.000	0.126	-0.193
ショートカットキーの表示	-0.173	0.356	0.126	1.000	-0.218
スキップの有効利用	-0.215	-0.272	-0.193	-0.218	1.000

表 5 各説明変数の相関係数 (glitch)

	解説量	メニューのズーム	マウスカーソルの強調	ショートカットキーの表示	スキップの有効利用
解説量	1.000	-0.268	-0.292	-0.558	-0.373
メニューのズーム	-0.268	1.000	0.242	0.454	0.032
マウスカーソルの表示	-0.292	0.242	1.000	0.302	-0.010
ショートカットキーの表示	-0.558	0.454	0.302	1.000	-0.034
スキップの有効利用	-0.373	0.032	-0.010	-0.034	1.000

表 6 各説明変数の相関係数 (crop + glitch)

	解説量	メニューのズーム	マウスカーソルの強調	ショートカットキーの表示	スキップの有効利用
解説量	1.000	-0.134	-0.191	-0.175	-0.251
メニューのズーム	-0.134	1.000	0.068	0.450	0.025
マウスカーソルの表示	-0.191	0.068	1.000	0.278	0.018
ショートカットキーの表示	-0.175	0.450	0.278	1.000	0.160
スキップの有効利用	-0.251	0.025	0.018	0.160	1.000

表 7 重回帰分析結果

説明変数	回帰係数 (crop)	回帰係数 (glitch)	回帰係数 (crop + glitch)
解説量	0.310	0.857	0.641
メニューのズーム	0.229	0.203	0.225
マウスカーソルの表示	0.035	0.116	0.077
スキップの有効利用	0.189	0.000	0.012
ショートカットキーの表示	-0.205	0.151	-0.158

4 関連研究

4.1 動画の編集, 構成技法

現在, 動画の編集技法や構成が, 視聴者の印象に及ぼす影響についての研究について活発に研究が行われている。

藤田ら [1] は, 動画の編集技法の一つであるトランジションに注目し, 視聴者に及ぼす影響について分析した。カットのトランジションを用いることで, インパクトや明晰さを視聴者に与えることができることを明らかにした。橋本ら [2] は, コンテンツの周囲において色刺激, CM, バナーの広告表示をすることで視聴者がどのような視線移動を行うかを分析した。色刺激について, 色の違いによってはあまり差がなく, 左右方向の視線移動が多いことを明らかにした。広告について, 視線移動の方向は個人差が大きく, バナーなどの静止画よりも CM などの動画の方が視線移動が多いことを明らかにした。Kim ら [3] は, 学生のオンライン講義の映像教材の視聴において, 動画の視聴の中断, 興味のピークのタイミングについて分析した。視聴の中断については, 指数関数的に中断する人数が増加することを明らかにした。興味のピークについては, 講義ベースの動画よりもチュートリアルベースの動画のほうがピーク数が多く, ス

ライドが変わったとき, スライドから操作が変わったとき, 講師が注目を促したときにピークが訪れることを明らかにした。津曲ら [4] は, レシピ動画において, 印象調査を行いわかりやすさの要因について分析した。規則性がわかりやすさの要因であることを明らかにした。

4.2 コンテンツ支援

ソフトウェア学習支援に関する研究も数多く行われており, 学習支援の形態として存在するコンテンツについて補足の支援を行うものと新たにコンテンツを生成するものが存在している。本節では前者を紹介し, 次節で後者を紹介する。

森下ら [5] は視線に着目し, 動画制作者の視線を記録し見どころとして扱い, 視線マーカーの軌跡を動画に重ね合わせることで特別な知識や技術, 人手による作業なしにコンテンツの見逃しを防止する研究を行った。坂本ら [6] はソフトウェア上での直前動作の取り消しの見逃しに着目し, 取り消しの動作を可視化, 通知, 内容提示を行うことで見逃しを防止する研究を行った。古田ら [7] は E-Learning 動画教材における動画視聴支援システムとして, 動画字幕から MMR を用いて自動でチャプターを生成した。Kim ら [8] は How-to 動画の視聴支援ツールのための研究として, クラウドソーシングで動画を操作ごとに分割

する研究を行っている。ワーカーに動画を分割させるだけではなく、結果をクラスタリングし、別タスクでその操作の解説とタイムスタンプが適切であるかどうかを確認することによって、有識者が分割を行った結果に近づけることを可能とした。

4.3 コンテンツ生成

Nguyen ら [9] はチュートリアル動画を作成する際、タイムライン上に操作を加えて手順を表示する動画を作成するツール DemoWiz [10] を用いて動画を作成し、画像処理によりユーザの操作と動画を同期させより短い時間でタスクを完了させることを可能とした。Chi ら [11] はソフトウェアを操作しログ、キャプチャ、入力デバイスのログを取得し、Web ページと動画を混ぜ合わせたチュートリアルを作成した。

5 おわりに

チュートリアル動画の理解の容易さを解説量と動画内で使用されている工夫から抽出する手法を提案した。解説量について、動画字幕から単語を抽出し、ソフトウェアで使用される特徴的な単語について χ^2 値を用いて重み付けを行い算出した。動画内の工夫について、人手でアノテーションを付与した。また動画の容易さについては、クラウドソーシング上で、Photoshop について知識があるワーカーに 2 つの動画についてどちらが学習に向いているかを質問し、回答を利用した。

算出した解説量と、クラウドソーシングで得られた動画内の工夫と動画の容易さについて重回帰分析を行った。

その結果、解説量が最も良い影響を及ぼすことが確認できた。解説量に続き、メニューのズームが理解容易性に良い影響を及ぼすことが確認できた。以上から、理解が容易なチュートリアル動画とは、ツールに対する解説が多く、メニューのズームがなされているものであると考えられる。ショートカットキーの表示については、画面上ではなく、音声で解説している動画が存在し、画面上の表示のみでは正しい結果を得られなかったため、ショートカットキーの説明の有無で分析を行う必要があると考えられる。

今後の課題として、楽器の演奏、ダンス等のソフトウェアでない他のジャンルのチュートリアル動画への適用、理解容易性因子が視聴者に与える影響を考慮したチュートリアル動画の提示を行っていく予定である。また、今回は線形回帰を用いて分析を行っているが、ランダムフォレスト回帰等、他の回帰分析も用いて、理解の容易さとなる因子を分析する手法も検討する。

謝 辞

本研究の一部は、平成 30 年度科研費基盤研究 (C)(課題番号: 18K11551) によるものです。ここに記して謝意を表すものとします。

文 献

- [1] 藤田良治, 山口由衣, 椎名健. 映像コンテンツの印象評価における編集技法の影響. 情報メディア研究, Vol. 7, No. 1, pp. 1–13, 2008.
- [2] 橋本圭輔, 牛木一成, 中村誠, 渡邊岳彦, 小河原成哲. 動画再生中における刺激提示の色の誘目性と配置に関する考察. 情報処理学会 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2006, No. 3, pp. 75–81, 2006.
- [3] Juho Kim, Philip J. Guo, Daniel T. Seaton, Piotr Mitros, Krzysztof Z. Gajos, and Robert C. Miller. Understanding In-video Dropouts and Interaction Peaks Inonline Lecture Videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pp. 31–40, 2014.
- [4] 津曲亜美, 杉山岳弘. SNS の視聴スタイルにおける魅力的な動画の要因分析とパターン化-レシピ動画の例から-. 第 80 回 情報処理学会 全国大会講演論文集, 第 2018 巻, pp. 783–784, 2018.
- [5] 森下浩平, 西田健志. 視線計測を利用した動画視聴支援システムの試作と評価実験, 日本認知科学会第 34 回大会. pp. P2–40, 2017.
- [6] 坂本有沙, 片山拓也, 寺田努, 塚本昌彦. デスクトップ上の画面変化に基づく取り消し操作の可視化機構の設計と実装. 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2012, No. 10, pp. 1–8, 2012.
- [7] 古田均, 白井理登, 中津功一朗, 高橋亨輔, 石橋健. E-learning における動画視聴支援システムの提案. 日本知能情報ファジィ学会ファジィシステムシンポジウム講演論文集, Vol. 28, pp. 643–648, 2012.
- [8] Juho Kim, Phu Tran Nguyen, Sarah Weir, Philip J. Guo, Robert C. Miller, and Krzysztof Z. Gajos. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pp. 4017–4026, 2014.
- [9] Cuong Nguyen and Feng Liu. Making Software Tutorial Video Responsive. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 1565–1568, New York, NY, USA, 2015. ACM.
- [10] Pei-Yu Chi, Bongshin Lee, and Steven Drucker. Demowiz: Re-performing software demonstrations for a live presentation. *Conference on Human Factors in Computing Systems - Proceedings*, 2014.
- [11] Pei-Yu Chi, Sally Ahn, Amanda Ren, Mira Dontcheva, Wilmot Li, and Björn Hartmann. Mixt: Automatic generation of step-by-step mixed media tutorials. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pp. 93–102, 2012.