

動画を活用した自己学習環境における文字起こしデータと視聴履歴からの自動理解度検出に向けて

齋藤 光貴^{1,a)} 渡辺 知恵美^{1,b)}

概要: 本稿では教材動画を用いた自己学習環境において字幕データと動画視聴履歴から学習者が理解に苦労した箇所の検出を行う仕組みを提案する。自己学習の流れとして、一通り動画を視聴したのち内容を学習者視点でまとめ、動画に対応する問題を解くという流れで行う。動画の文字起こしデータを用いて動画の内容を学習項目ごとにセグメンテーションし、それぞれの動画セグメントに対して理解度スコアを計算する。理解度スコアは自己学習の各フェーズにかかった作業時間と動画の視聴回数を合わせて計算することで、何度も見返し作業に苦心した箇所を検出する。また、大学数学の Youtube 動画を用いて予備実験を行い、動画視聴履歴の傾向を分析した結果、こまめな動画の一時停止や問題解答時の長時間の停止状態など特徴的な傾向を見つけることができた。

Toward Automatic Understanding Detection from Subtitle Data and View History in a Video-based Self-study Environment

1. はじめに

近年、オンライン学習環境が充実したことにより、自己学習の手段として動画を用いた学習が一般的になりつつある [2][3]。動画を用いた学習では、対面等での講義で理解が十分にできなかった箇所を繰り返し見直したり、逆にすでにわかっている箇所などを倍速で再生するなど、学習者の状況に合わせて柔軟に学習を進めることができる。さらに音声認識技術が大幅に向上したことにより、動画での発言内容を文字起こしし、字幕としてつけることができるようになった。Youtube では動画をアップロードすると自動的に音声認識を用いた文字起こしデータが作成される。音声認識の精度は動画内の雑音量や話し方、専門用語の数などに依存するが、講義動画では比較的精度の高い文字起こしデータが作成される。また、投稿者ツールによって文字の修正を行うことができる。

動画に対する字幕情報は動画の音声を再生することができない環境、聞こえないまたは聞き取りづらい環境などにおいて、動画の内容を理解することに役に立つ。その利点

に加えて、動画に付与された字幕データは発話内容を文字情報にすることができることから、主に以下の 2 点に活かすことができる。

- 発話内容を鳥瞰的に閲覧することができ、動画の構造を把握したり部分的な視聴ができる
- 自然言語処理を活用し、動画内容を機械的に構造化することができる

我々は、文字起こしデータが付与された動画における上記の特徴は特に教育系動画を用いた自己学習環境において最大限に活用できると考え、動画を活用した自習環境を提案する。我々が提案する自習環境の特徴は以下の通りである。

- 動画と発話内容の並列表記：
動画では講義内容を図などを用いてわかりやすく説明ができるという利点がある一方で、講義内容の概要を鳥瞰的に眺めることは難しい。一方で文字情報は内容を把握するのに時間がかかる可能性はあるが、1 ページに掲載される情報が多く、講義の内容や構造を掴みやすい。そこで、動画の横に文字起こしデータを並列表記することにより、動画とテキスト双方の利点を活かした学習ができる。
- 発話内容に基づく部分的な反復視聴

¹ 筑波技術大学
Tsukuba University of Technology
a) a211117@a.tsukuba-tech.ac.jp
b) chiemi@a.tsukuba-tech.ac.jp

講義内容の理解のための視聴では一通り視聴するだけでなく理解が十分でない箇所や確認をしたい箇所を何度も部分的に視聴することになる。文字起こしデータに発現時間の対応関係がついていると、部分的な視聴をスムーズに行うことができる。

- 部分的な反復視聴履歴を活用した理解度判定
部分的な反復視聴を繰り返すことにより、理解に時間のかかる箇所を重点的に学ぶことができる。またその履歴を記録することにより、学習者が理解に時間のかかった場所を特定することができ、学習者自身得意不得意を客観的に掴むことができる。

本稿では動画と文字起こしデータを併用した自己学習環境を提案し、その環境における動画視聴履歴から学習者の理解度を算出し可視化する手法について述べる。2節で関連研究を示したのち、3節では動画と文字起こしを活用した自己学習環境について現在開発中のプロトタイプシステムを用いて説明する。4節では提案する自己学習環境において動画視聴履歴を用いた学習項目別の理解度計算手法について述べ、そのランキング結果の評価を第5節に述べ第6節で考察をする。

2. 関連研究

学習履歴を活用した効率的な学習方法に関する研究は、教育データマイニング (Educational Data Mining: EDM) 及び学習分析 (Learning Analytics) 分野において数多く行われている。

その中で講義動画のクリックストリーム分析を用いた受講生の状況把握に関する研究も数多い。しかしながら、その研究の多くは対象とする Mooc のドロップアウト確率を予測するもの [4] や、ビデオに対する集中度の予測に着目するもの [7] などがある。動画視聴による受講者の知識の理解度をモデル化するとしては、Zhou らの研究 [8] がある。これはクリックストリームとテストの成績を学習し、クリックストリームから受講者のテストの成績を予測する試みである。単純にクリックストリームを使って予測した場合は実際のテスト結果との相関はほぼないという結果となっているが本研究では動画を用いた効率的な学習法と組み合わせているため、アプローチと効果は異なると考えている。

また、教育の実践現場においても教室で一斉に講義を行うのではなく、学習者個人の状況に合わせた個別進度学習を行うケースも増えてきている。また個別進度学習を ICT 技術でサポートする研究も行われている。[9] では個別進度学習で演習問題を解けない高原状態に陥った学習者を自動的に抽出するシステムを提案している。また、[10] では Web ベースのプログラミング学習授業支援システムにおいてクラス全体と問題のある学生の学習状況把握機能を提案し実装している。

3. 動画と文字起こしを活用した自己学習環境

3.1 想定する自己学習法

まず我々が動画を用いた効果的な自己学習方法として想定している学習の流れは以下の通りである。

- (1) 動画を一通り視聴する。途中で止めたり、早送りをしたり、見返したりするなど、動画の操作に関しては自由にして構わないが見ながらの内容の記録はメモ程度にとどめる。
- (2) 動画の内容を思い返ししながらその内容をノートにアウトプットする。ノートは前にとったメモとは別に書き出す。ノートはどんな媒体を使っても構わない。理解が曖昧な箇所や理解できていなかった箇所は、その旨をそのまま書き起こす。
- (3) ノートへのアウトプット時に理解が曖昧だった箇所に対する動画を見返して、ノートを完成させる。この際、動画は何度見返しても構わない。
- (4) 問題を解く。まずは何も見ずに問題を解くが、公式など、丸暗記をしなくて構わないものはメモとして傍において問題を解いても構わない。
- (5) 問題を解く際にわからなくなった箇所を動画を再度見直して確認をする。

この学習の流れは著者の経験則に基づくものであるが、以下の学習法とも関連する。

- Retrieval practice[6]: 参考資料を全く使わず、一度インプットした学習内容をアウトプットしようとすることで記憶の定着を促す。
- Brain dump: 直前に学んだことを一気に思い出してまとめることをいう。Retrieval practice の手法の一つ。我々が想定するノートへのアウトプットはこれに該当する。
- Hypercorrection effect[1]: 間違っていたことを修正されると、その項目は記憶に定着しやすいことを言う。ノートにまとめたり問題を解いたりして、曖昧だったり理解できなかったところを後で見返して修正するのはこの効果を狙っている。
- Test[5]: 練習問題を解き、問題を解くことで Brain dump とは異なるアウトプットを行い、記憶に定着させる。また答え合わせは Hypercorrection effect にも関連する。

3.2 動画と文字起こしデータによる学習

我々が想定する自己学習方法では、最初にインプットした内容を、繰り返しアウトプットしようと試み、アウトプットできなかった箇所を動画で見直し記憶の定着を測ると言うものである。この手法では、動画の部分的な箇所を何度も見返すこと、さらに見返したところが理解の定着の



図1 プロトタイプシステムのスクリーンショット

際に重要なファクターとなることがポイントとなる。

そこで、動画の全体的な構造を把握しやすくするために、動画と文字起こしのリストを並列表記する。図1にプロトタイプシステムのスクリーンショットを示す。

元プロトタイプでは動画はYoutubeに掲載されているコンテンツを利用することを想定している。Youtubeでは文字起こしデータは閲覧することはできるが、編集は投稿者にしか許されていない。現時点では音声認識システムにて自動文字起こしされたデータは編集なしでは十分に内容を理解するまでの精度には至っていない。そのため本システムでは一度文字起こしデータをYoutubeAPIを用いてSRT形式ファイルとしてダウンロードしたのち、手動で修正を行ったものを利用している。画面右側に文字起こし文章を表示し、文字の右側にあるリンクをクリックするとその文章を話している時間に動画をスキップさせることができる。また、動画の再生開始、終了時間を画面左下に記録している。

現時点では、文字起こしデータは時系列にただリスト表示されているに過ぎない。将来的には文章から動画を構造化し、その構造と学習項目を照らし合わせた形で表示できるようにしたいと考えている。ただし現時点の状況でも、講義内容をざっと眺め、動画の構造を理解するために文字起こしデータを利用することができる。

4. 動画視聴履歴を用いた理解度計算

我々は学習者の学習履歴から自動的に学習者の理解がまだ十分でない箇所、曖昧である学習項目の検出を行い、これを学習者に提示することで効率よく学習を進めていくことを目指す。そこで、学習履歴からの理解度計算を提案する。

4.1 動画の分割と学習項目への対応づけ

教材としての動画は、一つの動画の中で複数の学習項目が含まれている。我々は、意味のあるまとまりで動画をセグメンテーションしそれらを学習項目に当てはめることを検討している。ただし、現段階では手動でのセグメンテーションを行なっている。今後の方針として文字起こしデー

タを用いたセグメンテーションと、学習項目への関連付けを行う予定である。

4.2 セグメントごとの理解度計算

前述したように、我々の想定する自己学習手法では、自分の理解が曖昧であった箇所を動画で見返して確認をする。そのため、理解が曖昧または不十分である箇所は動画を見返す回数が増加すると考えることができる。また、まとめおよび問題の解答時間に関しても長くなると考えられる。これらのことから、動画の各セグメントに対して、動画の視聴回数および各フェーズに要する時間を用いて理解度のスコアリングを行う。

動画 V を n 個のセグメント $V = (v_1, \dots, v_n)$ に分けた時、セグメント v_i に対するユーザー u の理解度 $comp(u, v_i)$ の計算式は以下の通りとする。

$$comp(u, v_i) = w_0 + w_1 f_v(u, v_i) + w_2 f_s(u, v_i) + w_3 f_e(u, v_i) + w_4 d_s(u, v_i) + w_5 d_e(u, v_i)$$

ここで、 $f(u, v_i)$ は v_i を含む動画の視聴回数を表し、 f_v, f_s, f_e は学習のフェーズ (v :view=視聴, s :summerize=まとめ, e :exercise=練習問題) を表す。また、 $d(u, v_i)$ は v_i に対する作業時間 (duration) を表し、 d_s, d_e をそれぞれまとめにかかった時間、問題を解くのににかかった時間 (どちらも分単位) とする。

w_1, \dots, w_5 は視聴回数および作業時間 (分) に対する重みを表し、現段階では著者らによって設定をする。今後の展望としては、順序尺度に対する重回帰などを用いてこれらの重みを学習することを検討している。

5. 予備実験

5.1 目的と内容

本稿で提案している自主学習の方法で学習をした際の理解度計算について検証した。本実験の目的は以下の2点である。

- 学習内容の振り返りや問題解答を行う際、どのように動画を閲覧するか観察結果を得る
- 前節にて提案したセグメントごとの理解度計算は、実際の理解度にそくしているか

今回は上記の目的のうち1点目 (動画閲覧時の傾向を観察すること) を主目的とするため、予備実験として位置付ける。我々は「自己学習時に理解が不十分な箇所は、何度も確認するため閲覧回数が多くなる」ことを仮説として想定し、予想された行動が観測できるか、あるいはそれ以外の行動が観測できるかを検証する。

動画はYoutubeの講義動画を対象とし、統計確率における差の検定に関する講義*2 (前半:等分散の過程を置いた場合) を用いた。動画のセグメンテーションは著者の判断

*2 <https://www.youtube.com/watch?v=Det2IBRXajc>

表 1 まとめ及び問題解答にかかった時間

被験者	まとめ (秒)	問題解答 (秒)
A	68	308
B	126	878
C	327	1167

で以下の通りに設定した。

- 差の検定 (0:35-1:57)
- 等分散の仮定 (1:57-2:31)
- 確率分散 T の定理 (2:31-4:00)
- 例題の説明 (4:00- 5 : 00)
- 帰無仮説と対立仮説の設定 (5:00-5:29)
- t 値の計算と検定 (5:29-8:34)

また理解確認問題は該当する学習項目に合わせて1題用意した。被験者は情報系学部の2年生3名である。2名(被験者 A, B)が t 検定について講義で受講済みで1名(被験者 C)は受講していない。実験では3.1節にて述べた自己学習法に従って各自学習を進めてもらい、それぞれのフェーズにおける学習時間と動画の閲覧履歴を記録した。動画の閲覧履歴では、再生と一時停止した動画上での時間を記録した。式を確認するための動画閲覧が含まれにくくなるように、t 値の計算式と標準正規分布表を資料として被験者に提示した。実験実施後に各被験者にアンケートをとり、難しかった項目上位5つを記述してもらった。

5.2 実験結果

5.2.1 動画閲覧傾向の観察

被験者 ABC それぞれの動画閲覧履歴を分析した。まとめ及び問題解答にかかった時間を表1に示す。

- 被験者 A (講義動画内容の事前知識あり) :
一度動画を見た後、まとめおよび問題解答時には全く動画を閲覧することはなかった。これは被験者 A が事前に講義動画に該当する学習内容に関する十分な知識を持っており、手元の資料 (t 値の計算式と正規分布表) があればまとめや問題解答が完了できたことを表す。
- 被験者 B (講義動画内容の事前知識あり) :
一通り動画を閲覧したのち、まとめ時には動画を閲覧せず、問題解答時に以下の2回動画を閲覧したことが確認できた。
 - 0:00-2:58 差の検定の説明まで
 - 2:26-8:34 前置きを除き全体を一通り見直す (一時停止せず)

被験者 B も事前に講義内容の事前知識はあったため、何度も動画を見返すことはほぼなかった。問題解答時には、2:58 時点で一度一旦停止している。これは t 分布の定理について板書が確認できる部分であり、この時点で10分間停止しており、この定理 (特に自由度

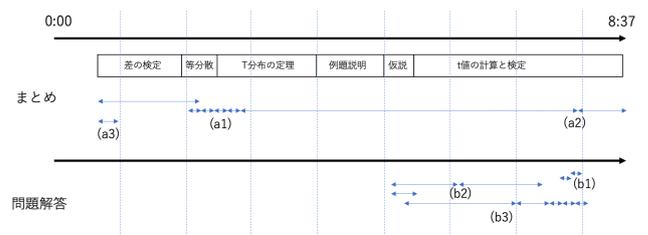


図 2 被験者 C の動画閲覧履歴

及び t 値を求める式) を閲覧しながら解答していた。その後説明部分を一通り見直しており、こちらは再確認のために一度問題をさらったと思われる。

- 被験者 C (講義動画内容の事前知識なし) :
被験者 C は講義動画に該当する学習内容は事前に学んでいないため、独学または予習での動画閲覧の状況を想定できる。被験者 C の動画閲覧履歴を図2に示す。動画の構成として前半に定理の説明、後半に例題の解説となっているため、まとめ時と問題解答時に閲覧する箇所が明確に異なっており、まとめ時は定理の説明部分を中心に閲覧し、問題解答時には例題解説部分を閲覧していることがわかる。まとめ時に一旦停止している部分 (図2(a1)) では t 分布の定理が全て板書されている部分、後半で一時停止している部分 (図2(a2)) は例題の回答が全て板書されている部分であり、板書の内容を中心に確認し書き写していた。また問題解答時に最初に一旦停止した部分 (図2(b1)) はまさに例題の板書部分であり、問題回答時に再確認したことが確認できる。しかしながらその後例題解説を2回繰り返して閲覧している (図2(b2), (b3))。特に2回目には t 値の計算から帰無仮説の棄却までこまめに一時停止をしていることがわかり、問題を解きながら確認をしていたことが見てとれる。

5.2.2 理解度の確認

本実験では1つの動画に対して問題は1問のみであり、まとめや回答時間に関してもどの学習項目に何分間かけていたかという詳細までは情報を取得していない。そのため今回は動画の各セグメントに対しての閲覧回数があるままランキングに反映されることとなる。表2に各被験者が動画を閲覧した回数、表3に各被験者が難しいと感じた項目を難しい順に並べたものを示す。被験者 B の閲覧回数を見ると「等分散の仮定」の部分での閲覧回数が最も多い一方で、アンケートでは自由度算出や棄却の判定に関する理解が難しいと答えている。動画閲覧履歴の分析においても述べているが「等分散の仮定」の説明箇所では被験者 B が動画を一旦停止をしているのは t 値の計算式を確認しているためであり、説明箇所を聞くためではない。実際、最も難しかったと答えた「自由度の計算」の理由について記号の意味の説明がなく理解するのに苦戦したと回答しており、その自由度の計算式が「等分散の仮定」の説明部分で板書画

表 2 動画を閲覧した回数

	被験者 A	被験者 B	被験者 C
差の検定	1	3	3
等分散の仮定	1	4	5
確率分散 T の定理	1	3	4
帰無仮説と対立仮説の設定	1	2	4
t 値の計算と検定	1	2	14

表 3 各被験者が難しいと感じた項目

被験者	難しいと感じた項目 (難しい順)
被験者 A	(なし)
被験者 B	自由度算出, 棄却の判定, t 分布上の t 値の位置 t 値の計算式, 仮説の設定
被験者 C	棄却の判定, 自由度算出, t 値の計算式 t 分布上の t 値の位置, 仮説の設定

面に表示されている。被験者 C の閲覧回数でも「等分散の仮定」の部分の閲覧回数が 5 回と多い。この理由も被験者 B と同様に板書を閲覧できる場所であることが観察結果より推察される。しかし圧倒的に多いのが「t 値の計算と検定」の 14 回である。アンケートを見ると最も難しかったのが棄却の判定、続いて自由度算出と t 値の計算となっており、これらの学習項目はすべて「t 値の計算と検定」に説明が含まれている。なお、自由度算出については動画内では受講者が既に理解していることを前提に講義されていたために動画内に明示的な説明がなく、被験者 B と C ともに回答時につまづいている。

6. 考察

今回の実験を通して、以下の観察結果及び知見が得られた。

- 講義動画には基礎知識の説明部分と例題を用いた説明部分があり、必ずしも動画のセグメントと学習項目が一致するわけではない
- 部分的な動画の閲覧目的は理解が不十分である箇所の確認のほかに、板書内容の確認もあり、学習者が理解したい部分が必ずしも口頭での説明部分と一致するわけではない
- 板書内容を確認する適切な場所をシークするために細かく動画の一時停止とスキップを行う場合もある
- 板書内容をもとに問題を解答する際には次の再生まで間があいている。
- 問題の解答時に例題を細かく一旦停止しながら閲覧する場合は動画の通りにひとまず問題を解いてみるという行動が観察された

1 点目は今回著者により手動でセグメンテーションをしたが、将来的に自動でセグメンテーションをする際にかなり技術的にチャレンジングな問題になると感じた。2 点目に関しても動画の再生履歴だけでは抽出できず画像処理等が必要になる難しい部分となる。3 点目と 4 点目に関しては、

動画の一旦停止と再開のパターンが特徴的であるため、自動的に抽出して理解度判定で考慮に入れるためのヒントとなり得る。5 点目に関しては理解度計算をする上での大きなヒントとなり得る。つまりこまめに一旦停止をし、次の再生までに数分間の細かい間がある場合は動画を見ながら問題を解いている可能性がある。そしてそのような場合、学習項目を十分に理解できていない可能性が示唆される。今回は事例が少ないため更なる検証が必要となるが、大きなヒントとなり得る。

また、今回被験者 A、B は復習として今回の学習に取り組み、被験者 C は独学として今回の学習に取り組んでいる。そのため元々の理解度にも差があったが、自己学習の目的が何であるかによって、学習の流れや動画の閲覧パターンも異なってくるのではないかと考えられる。例えば被験者 C は学習内容を確認するためにこまめに一旦停止をして学習をしていたが、被験者 A、B は全体を流して閲覧をするのがメインで、以前に学習した内容の振り返りが主な作業内容となってくるかもしれない。このような目的の違いも考慮に含めることも考えられる。

7. まとめと今後の課題

講義動画と文字起こしデータを用いた効率的な自己学習環境の提案をおこなった。文字起こしデータを用いて講義動画の構造を把握し、何度も試聴したい箇所を部分的に見直すことで理解の定着を促すことを目的とし、アウトプットを主体とした動画視聴スタイルによる学習法と併せて提案した。また、動画の閲覧履歴から理解度検出を試み、予備実験を行った。予備実験の結果、板書の確認を目的とした視聴など試聴履歴データだけでは検出できない行動がある一方、何度も動画を一旦停止して視聴する場合は理解が十分でないのではないかと仮説を立てることができた。

今回は予備実験ということで本格的な検証を行っていないため、今後の知見をもとに理解度検出の本格的な案と検出を進めていきたい。また、文字起こしデータと動画の併用によって動画の全体的な構造把握と部分視聴が促進されているかどうかの検証を行っていききたいと考えている。

参考文献

- [1] Butterfield, B. and Metcalfe, J.: *Errors committed with high confidence are hypercorrected*, Journal of Experimental Psychology: Learning, Memory and Cognition, 27(6), pp.1491-1494 (2001).
- [2] Coursera: 入手先 (<https://www.coursera.org/>), (2022.12.5)
- [3] edX: 入手先 (www.edx.org), (2022.12.5)
- [4] Ahmed A Mubarak, Han Cao, and Weizhen Zhang: Prediction of students' early dropout based on their interaction logs in online learning environment, Interactive Learning Environments, pp.1-20 (2020).

- [5] Rickard, T. C. and Pan, S. C.: *A dual memory theory of the testing effect*, *Psychonomic Bulletin and Review*, pp.1-23 (2017).
- [6] Roediger III, H. L. and Butler, A. C.: *The critical role of retrieval practice in long-term retention. Trends in Cognitive Sciences*, 15(1), pp.20-27 (2011).
- [7] Tanmay Sinha, Nan Li, Patrick Jermann, and Pierre Dillenbourg: Capturing “attrition intensifying” structural traits from didactic interaction sequences of MOOC learners. In Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs, pp.42-49 (2014).
- [8] Guojing Zhou, Tetsumichi Umada, and Sidney D’Mello: What do Students’ Interactions with Online Lecture Videos Reveal about their Learning?, In Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP ’22), pp.295-305 (2022).
- [9] 桑原 恒夫, 玉城 幹介, 山田 光一, 中村 喜宏, 西沢 正樹, 小西 納子, 天野 和哉, 弘田 健二 and 近藤 正紀: 個人進捗別教育支援システム (MESIA) におけるテストと小テストの分離による認知的効果, 電子情報通信学会技術研究報告. ET, 教育工学, 99(161), pp.39-45 (1999).
- [10] 加藤 利康 and 石川 孝: プログラミング演習のための授業支援システムにおける学習状況把握機能の実現, 情報処理学会論文誌, 55(8), pp.1918-1930 (2014).