

ビデオ講義における学習者の視線情報を用いた理解度の予測

村田 大輝[†]
兵庫県立大学[†]

川嶋 宏彰[‡]
兵庫県立大学[‡]

1 視線情報を用いた理解度の予測

視線は、ビデオ講義における学習者の状態を把握するための重要な情報である。視線から抽出される様々な特徴量は、受講中の mind wandering をはじめとする学習者の注意状態の推定 [1, 2] に利用できるだけでなく、テストの得点との有意な相関が確認されている [3]。学習内容ごとの詳細な理解度の推定は、効果的な学習支援や教材の改良につながる。そこで本研究は、ビデオ講義視聴中の学習者の視線情報から、その学習者の理解度、具体的には講義に関する各問題での正誤を予測することを目的とする。

本研究の特徴は2点ある。1つは、視線特徴量と理解度の関係分析 [3] にとどまらず、講義内容の間に対する正誤予測を直接的に行う点、もう1つは学習者の「注意状態」を考慮する点である。例えば、学習者が講師の話の内容とは関係のない部分を見ている場合には、「スライドを先読みしている」、「講師やスライド以外のことに注意を向けている」などの注意状態が考えられる。抽出した視線特徴量だけでなく、このような注意状態による影響を考慮することで、より正確な理解度予測につながる可能性がある。そこで、「学習者の視線の動きは、学習者の理解度だけでなく学習者の注意状態にも依存している」という仮説に基づき、視線特徴量に加え、視線データから推定した注意状態の特徴を用いた際の理解度推定の精度について比較・検証を行う。

2 視線特徴の抽出と注意状態の推定

2.1 全体の流れ

講義ビデオは講師がスライドを用いて説明する形式を想定する。画面上のコンテンツ（講師、テキスト、表など）の領域を抽出し、AOI (area-of-interest) として設定する。一方、視線データからは注視点系列を検出し、ここから注視領域系列を得る。これらの情報から抽出した視線特徴量および推定した注意状態の特徴量を用いて、学習者の理解度を推定する。

2.2 視線特徴量

視線特徴量として、各スライドで以下の2種類の特徴量を抽出する。なお、注視点の抽出には頑健性に優れた「分散閾値による検出」 [4] を採用した。

コンテンツ非依存の特徴量： スライドが開始してから次のスライドに切り替わるまでの注視の回数、各注視の継続時間の記述統計量（平均、標準偏差、最小値、中央値、最大値など）、各注視点の位置 (x 座標, y 座標) の記述統計量、視線が画面外に出た時間の割合

コンテンツ依存の特徴量： 各 AOI 内に注視点が入る回数と時間（合計と割合）、各タイプ（講師、スライドのタイトル、テキスト、赤字を含む強調テキスト、表）の AOI に注視点が入る回数と時間（合計と割合）、いずれかの AOI に注視点が入る回数と時間（合計と割合）

2.3 注意状態特徴量

注意のモードを潜在変数とした確率的生成モデルを用いて、注視領域系列から学習者の注意状態の推定や視聴スタイルの特徴づけを行う手法が提案されている [2]。本研究ではこの提案モデルを用いて注視領域系列から学習者の注意状態を推定する。

入力とする注視領域系列の時間区間にはスライド切り替えが含まれないものとする。画面中の AOI の ID 集合を $\mathcal{R} := \{R_1, \dots, R_N\}$ とし、学習者の k 番目の注視領域を $r_k \in \mathcal{R}$ とすると、注視領域系列は $r_1, \dots, r_{k-1}, r_k, \dots, r_K$ と表される。また、 r_k を注視した時刻（スライドの開始時刻に対する相対時刻）を t_k 、そのときの注意状態を $m_k \in \{0, 1, 2\}$ とする。

ビデオ講義視聴中の学習者の注意状態として、以下の3種類のモードが存在すると仮定する（式は各モードにおける注視領域の分布モデル）。

モード 0： mind wandering など、スライドと講師のガイド以外に注意している状態。

$$P(r_k = R_i | m_k = 0) = a_i$$

モード 1： スライドの意味内容を積極的に追う状態。 $P(r_k = R_j | r_{k-1} = R_i, m_k = 1) = b_{ij}$

モード 2： 講師のガイド（発話や指示棒など）を追う状態。 $P(r_k = R_i | t_k = t, m_k = 2) = c_{it}$

Predicting Comprehension Using Learners' Gaze Information in Video Lectures

[†] Taiki Murata, University of Hyogo

[‡] Hiroaki Kawashima, University of Hyogo

さらに隠れマルコフモデル (HMM) と同様に、注意状態はモード遷移確率 $P(m_k = q | m_{k-1} = p) = A_{pq}$ ($p, q \in \{0, 1, 2\}$) によって確率的に切り替わるとする。すなわち、学習者の注意状態が時間的に遷移するとともに、どの注意状態にあるかで注意領域の分布が切り替わると仮定した、注視領域系列の生成モデルである。逆に注視領域系列が与えられれば、モデルの学習と共に、学習者の注意状態の系列 m_1, \dots, m_K を最尤推定できる。

注意状態特徴量 推定された注意状態系列より、各モードの出現回数と出現時間 (合計と割合)、各モード遷移の回数 (合計と割合) を注意状態特徴量として抽出する。出現時間は t_k と t_{k-1} の差を用いる。さらに、これらの特徴量に時刻による重み付けを行うことで、時刻を考慮した特徴量を抽出する。出現回数では時刻を k として重み付けを行い、出現時間では時刻を t_k として重み付けを行う。早い時刻に大きな重みを付けた特徴量と、遅い時刻に大きな重みを付けた特徴量を抽出する。重み w_k の範囲は、 $0 \leq w_k \leq 1$ とする。

3 実験

視線特徴量及び注意状態特徴量と、学習者の理解度の関係を分析するために、講義ビデオ視聴後のテストの問題の結果と、その問題に主に対応するスライドでの視線特徴量及び注意状態特徴量を用いた実験を行った。

3.1 データセット

1本の講義ビデオと、その講義ビデオを視聴した28名の視線データを用いた。講義ビデオは、約14分の長さで、音声やスライド画像、講師 (動き、指示棒) などのデータが含まれている。「統計的推定」をテーマとした講義内容となっている。理解度を表す指標として、講義動画視聴後に実施したテストの結果を用いた。テストは正誤選択式の問題と、4つの選択肢から1つを選択する穴埋め問題の2種類からなる。13問のテストのうち、正解者と不正解者の比率の偏りが比較的小さな問題 (問2, 11, 9, 5) を選択した。

3.2 特徴量の有意差に関する統計解析

各問題に主に対応するスライドを選び、問題の正解者群と不正解者群の間で、そのスライドにおけるどの視線特徴量や注意状態特徴量に有意差があるかを、Welchのt検定 (有意水準5%) を用いて調べた。その結果、問題によっては、注視点の位置や、視線が画面外に出た時間の割合などのコンテンツ非依存の視線特徴量、スライドのタイトル領域や、問題の内容に関わるAOIなどに関するコンテンツ依存の視線特徴量、モード0に関する注意状態特徴量などに有意な差が確認された。

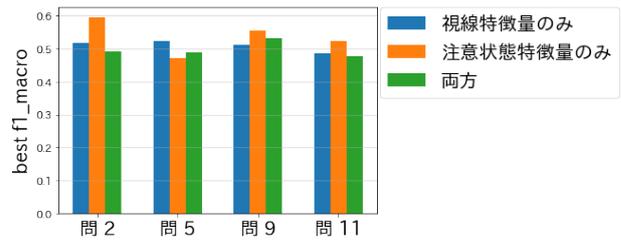


図1 予測精度 (問2, 5, 9, 11)

3.3 理解度の予測

特徴量の統計的パターンを機械学習し、学習者の理解度を予測した。具体的には、勾配ブースティング (実装はLightGBM) を用いて、問題に主に対応するスライドにおける特徴量を用い、その問題に学習者が正解するかどうかを予測した。評価指標をF1マクロ平均とし、leave-one-out交差検証により、視線特徴量のみを用いた場合、注意状態特徴量のみを用いた場合、両方の特徴量を用いた場合の予測精度の比較を行った。LightGBMのハイパーパラメータは、視線特徴量と注意状態特徴量の両方用いて、F1マクロ平均を最大化するように各問で最適化を行った。さらに、正解者数と不正解者数の比率の偏りを考慮するために、学習時の勾配に重みづけを行った。

図1に、各問題における予測精度 (F1マクロ平均) を示す。乱数シードを20回変化させて行った予測の精度の平均を、いくつかのブースティング反復回数で計算し、最も高い精度を示した反復回数における予測精度を採用した。問2, 9, 11では、注意状態特徴量のみを用いた場合の予測精度は、視線特徴量のみを用いた場合よりも高いことが確認された。これより、注意状態が学習者の理解度に関わっていることが示唆された。

謝辞 本研究の一部は科研費JP19H04226の補助を受けて行った。

参考文献

- [1] Stephen Hutt, Jessica Hardey, Robert Bixler, Angela Stewart, Evan Risko, and Sidney K. D'Mello. Gaze-based detection of mind wandering during lecture viewing. *Proc. of the International Conference on Educational Data Mining*, pp. 226–231, 2017.
- [2] Hiroaki Kawashima, Kousuke Ueki, and Kei Shimonishi. Modeling video viewing styles with probabilistic mode switching. *Proc. of the International Conference on Computers in Education*, pp. 81–86, 2019.
- [3] Kshitij Sharma, Patrick Jermann, and Pierre Dillenbourg. “with-me-ness”: A gaze-measure for students’ attention in MOOCs. *Proc. of International Conference of the Learning Sciences*, pp. 1017–1022, 2014.
- [4] Dario D. Salvucci and Joseph H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. *Proc. of the Symposium on Eye Tracking Research & Applications*, pp. 71–78, 2000.