

講義室での受講生の振る舞い観測と理解度推定の研究

Observation of Students' Behavior and Estimation of their Comprehension Level in a Classroom

棕木 雅之 美濃 導彦
Masayuki MUKUNOKI Michihiko MINOH

京都大学学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University

In this paper, we introduce our research on analyzing the relationship between student's behavior and his comprehension level in a classroom. We regard lectures as communication between an instructor and students and we evaluate lectures from a viewpoint whether the information has reliably transmitted between them. For that purpose, we develop two tools. One is a tool to observe the students' behavior from a students' video footage, which reduces the cost for the analysis, and the other one is a tool to estimate the student's comprehension level using the questionnaire about his objective comprehension level and the quiz on the contents of the lecture. We also show a preliminary result on categorizing the students into typical groups based on their behavior and analyzing each group in terms of their behavior and comprehension level.

1. まえがき

近年、我が国では、高等教育におけるFD (Faculty Development) が広がりを見せている。FDの一環として講義評価は多くの大学で実施されており、FDの中でも重要な位置づけにあると考えられる。講義評価に関して、講義の様子を撮影した講義映像を観察して講義を評価する取り組みはすでに行われているが、その多くでは講師の教示法に注目している。しかし、我々は講義を、講師と受講者の双方向コミュニケーションと捉えており、講義評価においては、講師の教示に対する受講者の反応にも注目する必要があると考えている。

一斉講義においては、受講者の反応は主に振る舞いとして表れる。受講者の振る舞いは、外部から観測可能であるが、受講者の振る舞いと講義の善し悪しの関係性は自明ではない。一方で、講義評価の指標として受講者の理解度も挙げられる。受講者の理解度が高い講義は、情報伝達が適切に行えた良い講義とみなすことができるが、理解度は受講者の内面の変化であり観測は難しい。

そこで我々は、講義における受講者の振る舞いと理解度の関係性を明らかにすることを目指して、研究を行っている。両者の関係性が明らかになれば、受講者の振る舞いから講義を評価することが可能になり有用である。これを実現するには、「受講者の振る舞いの観測」「受講者の理解度の推定」「振る舞いと理解度の関連性の解析」という3つの技術要素が必要となる。本稿では、これらの技術要素について我々がこれまで取り組んできた研究成果について述べる。

2. 時空間制約を用いた受講者の姿勢系列推定

2.1 振る舞いと姿勢系列

従来から、講義中の受講者の振る舞いに着目して講義を分析する研究が行われている。米谷ら [1] は、受講者の「見る行動」「上半身を支える行動」「手の動作」などの振る舞いに着目して、講義の分析を行っている。一方、溝上ら [2] は、受講者の「顔上げ」に着目した分析を行っている。このように、講義

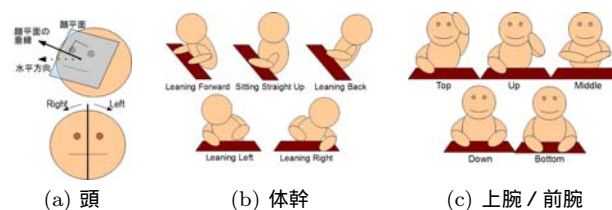


図 1: 部位状態

中の受講者の振る舞いとして何に着目するかは、どのような分析を行うかに依存しており、一定していない。

また、これらの研究では、90分程度の講義映像を人間が見て、受講者の振る舞いを書き出している。これは、非常に負担のかかる作業であり、計算機により自動化することが望まれる。計算機により自動推定することを考えると、様々な振る舞い分類に汎用的に適用できるツールとして技術開発することが望ましく、受講者の振る舞いを網羅的に記述できる方法が必要である。

これらの点を考慮して、本研究では、受講者の各瞬間での姿勢を時間方向に並べた姿勢系列を推定するアプローチをとる。姿勢は、頭、体幹、上腕/前腕の部位の3次元位置・向きを組み合わせにより表現する。これらは、体の部位の3次元位置という定量的な基準に基づく分類であり、客観性・網羅性が高い。また、振る舞いは、姿勢系列の特定の性質をもつ部分に意味づけを与えることで得られるため、姿勢系列から比較的容易に推定することができる。一旦、受講生映像から姿勢系列が得られれば、受講生の振る舞い分析の着目点が変わっても、その振る舞いの定義に応じた姿勢系列を抽出することで対応可能であることから、姿勢系列は汎用性が高い。

具体的には、図1に示すように、頭4状態×体幹5状態×上腕/前腕5状態 = 100通りの組み合わせにより、姿勢を表現する。

2.2 姿勢系列推定の問題点

人体の姿勢を推定する手法には、大きく分けて、推定対象となる人物にマーカを取り付ける手法と、カメラ等による外部からの観測のみを利用する手法がある。この内、マーカを取

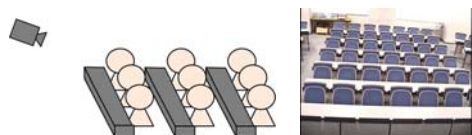


図 2: 受講者映像のカメラ配置



図 3: Ferrari らの手法の結果例

り付ける手法は、数十人の受講者に適用するには、費用、設備面で難しいだけでなく、装着により受講者に違和感を与えてしまうことから、講義室での利用には不適である。外部からの観測のみによる手法として、例えば Kinect[3] 等の距離センサを用いることも考えられるが、計測範囲が狭いことから、講義室への適用は不向きである。本研究では、図 2 に示したように、講義室前上方に設置したカメラで受講者を撮影した映像から、カメラ内の受講者の姿勢を推定することを目指す。

カメラ映像内の人物姿勢を推定する研究は、いくつか行われている。例えば、Ferrari ら [4] は、画像のエッジ情報を利用して、人体の部位を検出し、姿勢推定を行っている。この手法は、画像内に複数の人物が存在しても適用でき、輪郭を用いる手法に比べ、後ろに別の人物が重なっていても処理可能であるなど、受講者映像への適用に有利な特徴を持つ。一方で、受講者映像では、受講者の部位間の重なりが多く、姿勢推定に失敗することが多いという問題点もある。図 3 は、Ferrari の手法を受講者映像に適用した例である。色のついた線分が各部位の検出結果である。体幹と重なっている上腕 / 前腕の検出に失敗する例が多くみられ、この部分の改善が必要である。

2.3 時空間制約の導入

本研究では、Ferrari らの手法を基に、姿勢の時空間制約を導入することで、姿勢推定の精度向上を計る。講義室において、受講者は座席に座っているため、講義中にとる姿勢はある程度限られている。例えば、座席と机の位置関係から、体幹を後ろに傾けたまま、上腕 / 前腕を机の上に置くような姿勢を取ることは少ない。このような部位間の空間的關係を「姿勢モデル」として導入し、講義室で現れやすい姿勢を推定結果としてより優先させる。同様に、時間方向に姿勢が急に変化することも稀であるため、姿勢間の時間的な変化の關係を「遷移モデル」として導入し、より現れやすい姿勢遷移を推定結果として優先させる。このように、時空間制約を導入することにより、講義室において表出されやすい姿勢を優先して、姿勢推定結果を補正し時間方向に平滑化することで、推定精度の向上を計る。

具体的には、以下の手順で姿勢推定を行う。

1. Ferrari らの手法により部位を検出
2. 「姿勢モデル」を用いて、部位の相互關係から部位状態の信頼度を算出
3. 「遷移モデル」を用いて、時間方向に信頼度のある時間幅 K の区間で伝播することで、姿勢の確信度を算出
4. 映像の各時刻について、1. から 3. を時刻順に適用
5. 最終的に各時刻で最も確信度の高い姿勢を、推定結果とする

姿勢モデルは、ある部位 n がある状態 s_n である時に、別の部位 m がある状態 s_m である条件付き確率 $P(s_m|s_n)$ で表す。部位の検出位置から算出した初期的な部位状態の評価値を C_{parts} とすると、姿勢モデルを適用した後の部位状態の信頼度 C_{pos} を次式で求める。

$$C_{pos}(s_m, t) = \sum_{s_n \in \{s_{head}, s_{torso}, s_{hands}\}} P(s_m|s_n) C_{parts}(s_n, t).$$

遷移モデルは、ある時刻 t での姿勢 λ_p^t とその前の時刻 $t-1$ での姿勢 λ_q^{t-1} の間の条件付き確率 $P(\lambda_p^t|\lambda_q^{t-1})$, $P(\lambda_q^{t-1}|\lambda_p^t)$ で表す。ある時刻 t での姿勢 $\lambda^t = (s_{head}, s_{torso}, s_{hands})$ の初期確信度 $B_{pos}(\lambda^t)$ は、その姿勢を構成する部位状態の信頼度の積で計算する。

$$B_{pos}(\lambda, t) = C_{pos}(s_{arm}, t) C_{pos}(s_{torso}, t) C_{pos}(s_{head}, t).$$

この初期確信度を、時間軸を遡る方向に時間区間 K の範囲で伝播した後、逆に時間軸方向に伝播させることで、時間軸方向の姿勢の補正・平滑化を行う。

$$B_{rev}(\lambda_q, t-k-1) = \sum_{\lambda_p \in \Lambda} P(\lambda_q^{t-1}|\lambda_p^t) B_{rev}(\lambda_p, t-k)$$

$$B_{trans}(\lambda_p, t-k) = \sum_{\lambda_q \in \Lambda} P(\lambda_p^t|\lambda_q^{t-1}) B_{trans}(\lambda_q, t-k-1)$$

2.4 実験結果

時空間制約の導入による姿勢推定の精度向上について、実験を行った。実験では、図 2 のカメラ映像中の、前 2 列の座席の受講者 7 人について、3 分間 1800 フレームの映像から姿勢推定を行った。姿勢モデルとしては、以下の 3 つの關係が現れやすいとした。

- 上腕 / 前腕を下におろし (Down)、体幹を前に傾ける (Lean-Forward)
- 上腕 / 前腕を机に載せ (Bottom)、体幹を前に傾ける (Lean-Forward)
- 上腕 / 前腕を机から上げ (Middle)、体幹を後ろ傾ける (Lean-Back)

遷移モデルとしては、以下の 2 つの關係が現れやすいとした。

- 同じ姿勢への遷移
- 上腕 / 前腕および体幹の状態は同じで、頭のみ上げ下げ (Up/Down) する遷移

これらのモデルは、受講者映像を観察して人手で与えた。また、それぞれの条件付き確率も、実験的に定めた。

目視で与えた部位状態、姿勢の正解との一致度を表 1 に示す「あり」が本研究での結果、「なし」は時空間制約を用いなかった場合の結果である。

全体的に見て、モデルを適用することで精度が向上している。受講者 A・B に関しては、著しい精度向上が見られた。また、受講者 C・D・E・G も一定の精度向上がみられ、受講者のとりやすい姿勢に基づく制約を与える本研究の方法は、様々な受講者に対して適用できるといえる。受講者 F に関しては、姿勢モデル、遷移モデルとして与えた、受講者がとりやすい姿勢として想定していたものとは異なる姿勢をとっていたため、上腕 / 前腕の部位状態の精度が低下している。ただし、本研究

表 1: 各受講者の部位状態ごとの正解率と姿勢の正解率 (%)

	上腕/前腕		体幹		顔		姿勢	
	なし	あり	なし	あり	なし	あり	なし	あり
受講者 A	5.7	67.3	9.1	59.7	53.2	53.5	0.8	38.4
受講者 B	34.7	80.2	11.7	98.8	62.6	62.3	1.6	52.5
受講者 C	92.9	92.6	37.1	46.9	77.6	78.1	20.1	34.0
受講者 D	41.0	51.7	27.8	52.5	72.3	72.4	6.2	15.0
受講者 E	12.2	13.9	37.1	55.3	79.9	80.2	0.1	12.3
受講者 F	30.5	19.0	38.8	69.6	85.5	85.8	0.0	10.1
受講者 G	18.2	49.2	17.3	76.9	90.5	91.1	8.6	26.1

では制約に無理矢理あてはめることは行っていないため、上腕/前腕の正解率は 0% となることはなく、他の体幹や頭の正解率は向上が見られ、制約による負の影響が低く抑えられていることがわかる。

姿勢全体としては、正解率は高くないものの、各部位の正解率の最小値から大きく低下することはなかった。姿勢は部位の組み合わせであり、部位同士を独立に扱おうと、姿勢の正解率は各部位の正解率の乗算ほどに低下する。しかし、表 1 の姿勢の「あり」の推定結果は、各部位の正解率の乗算ほどは低下しておらず、本研究で提案した手法の適用により、部位状態の間の関係の制約が作用し、姿勢の正解率の低下がおさえられていることがわかる。

3. 項目反応理論を用いた受講者の理解度推定

3.1 受講者の理解度推定

受講者の振る舞いと理解度の関係を解析するためには、あらかじめ受講者の理解度を推定しておく必要がある。理解度を推定する手法としては、インタビューによる方法や、多数の問題を出題しその正解率をみる方法、講義の前後でテストを実施しその成績の変化をみる方法など、種々の方法が考えられるが、実際の講義で実施するには、講義の進行を妨げず、受講者に負担とならない方法が必要である。本研究では、受講者自身に理解度を申告してもらう理解度アンケートによる方法と、問題を出題しその正解率を用いる小テストによる方法を併用する。講義を数個のスパンに分け、スパン毎に理解度アンケートと小テストを実施する。スパンとして 30 分程度の比較的短い期間単位に講義を区切ること、受講者の振る舞いと対応付けやすい単位での理解度推定を目指す。

理解度アンケートは、受講者自身が講義内容をどれほど理解していると思っているか、すなわち、受講者の意識による主観的な理解度を測る指標である。アンケート形式は五件法とし、選択肢は以下の五つとする。

非常によく理解できた	100 ~ 90%
よく理解できた	90 ~ 80%
だいたい理解できた	80 ~ 70%
まあまあ理解できた	70 ~ 60%
あまり理解できなかった	60 ~ 0%

「非常によく理解できた」等の記述のみでは、講義内容をどれだけ理解できたときにその選択肢を選ぶかの判断基準が個人によって異なると考えられるため、本研究では、選択肢の理解度をパーセンテージを用いて表した。

小テストは、出題した問題に対する正解率により理解度を測る指標である。小テストの形式は、直前のスパンでの講義内容に関する記述を提示し、その正誤を問う正誤問題を 5 問出題する。5 問中の正解した割合を、正解率とする。提示する講義内容に関する記述は、教材に含まれる内容の他に、講師が板書

や口頭説明した内容、それらから類推できる内容を含む。受講者は「正しい」「誤り」「わからない」の 3 つの選択肢から回答を選ぶ。「わからない」の選択肢は、解答者が提示された記述の正誤が判断できない場合に、勘のみで解答し偶然に問題に正解する事態を防ぐために用意した。解答者が「わからない」と解答した場合は、不正解として扱う。

3.2 個人差と難易度差

理解度アンケートや小テストは、受講者の理解度を測る指標ではあるが、そのままでは、異なる受講者間や異なるスパン間での理解度を比較する際の基準にはならない。

実際に、異なる受講者が理解度アンケートに同じ回答をしても、受講者によって小テストの正解率が大きく異なる傾向があった。これは、主観的な理解度の解釈に個人差があることを示している。本研究では、理解度アンケートの選択肢として具体的な数値を併記することで、ある程度絶対的な尺度に沿った回答が得られることを期待したが、それ以上に個人差が大きく現れるという結果になった。このため、受講者ごとに理解度アンケート回答の個人差を調整しなければ、異なる受講者の回答した理解度アンケートを同一の尺度で評価することができない。

また、同じ受講者が理解度アンケートに対して同じ回答をした場合でも、小テストの正解率が大きく異なる場合があった。これは、小テストの問題に難易度差があったためと考えられる。問題作成時には、小テストの難易度が均等になるよう配慮したが、小テストを解答した受講者への聞き取りでは、小テストによって難易度が異なるという意見が寄せられた。小テスト間で難易度が等しくなければ、正解率の違いが小テストの難易度によるものなのか解答した受講者の理解度によるものなのかを判断できず、受講者の理解度を測る一定の評価基準になり得ない。

3.3 項目反応理論の導入

異なる難易度のテストを用いて解答者の能力を測る際に用いられる理論の一つに、項目反応理論 [5] がある。項目反応理論とは教育学分野における「試験」に関する理論であり、問題の正答率 p と解答者の能力 θ の関係が次式のロジスティック曲線で表現されるとするものである。

$$p(\theta) = c + \frac{1 - c}{1 + e^{-1.701a(\theta - b)}} \quad (1)$$

a, b, c はパラメータであり、意味は以下の通りである。

- a : 識別パラメータ。問題が解答者の能力を識別する力を表す実数値
- b : 難易度パラメータ。問題の難しさを表す実数値であり、一般的には、50% の正解率を持つ解答者の能力値を基準として決められる
- c : 多肢選択形式のテストの場合に、解答者が偶然に正解できる確率を表す実数値 (本研究では $c = 0$ とする)。

項目反応理論は、解答者やテストの難易度によらず共通の尺度で解答者の能力 θ を測定できる。小山ら [6] は、解答者の能力 θ を客観的かつ絶対評価の理解度と記述している。本研究において理解度アンケートによって収集する主観的理解度は、 θ とは必ずしも一致しないが、「受講者が講義内容をよく理解したと思った際に客観的な理解度が高い」というように、相関があると考えられる。

スパン j に対応する理解度アンケートにおける受講者 i の回答を ψ_{ij} 、このスパンの内容に対する受講者の能力 (客観的

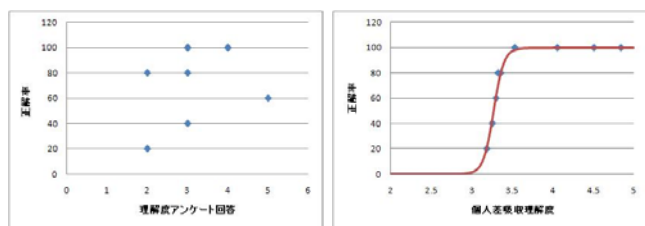


図 4: 補正前

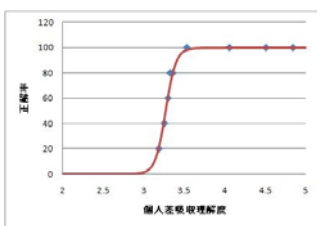


図 5: 補正後

理解度)を θ'_{ij} として、両者の間に以下の関係が成り立つと仮定する。

$$\theta'_{ij} = x_i + y_i(\psi_{ij} - 1) \quad (2)$$

ここで、 x_i は受講者 i が理解度アンケートに「1:あまり理解できなかった」と回答した際の基準理解度であり、 y_i は受講者 i の理解度アンケート回答における各段階の基準理解度の間隔である。

本研究では、上式で得られる客観的理解度と小テストの成績との間に項目反応理論が成り立つと仮定し、式(1),(2)のパラメータを求めることで、理解度アンケート回答の個人差、小テストの難易度差を補正する。

データ収集対象とする受講者の数を n 人、実施した理解度アンケート・小テストの数を m 回とする。小テスト j の従うロジスティック曲線の識別パラメータを a_j 、難易度パラメータを b_j とすると、受講者 i の小テスト j における正解率の予測値 p_{ij} は、式(1)で求められる。受講者 i の小テスト j における正解率の実測値を P_{ij} とすると、 m 回すべての小テストにおける正解率の予測値と実測値の二乗誤差 E は次式で与えられる。

$$E(\theta', a, b) = \sum_{j=1}^m \sum_{i=1}^n (P_{ij} - p_{ij})^2 \quad (3)$$

E の値が最小になるようにパラメータ x_i, y_i, a_j, b_j を決定することで、客観的理解度を求める。

個人差、難易度差の補正の例を図4, 5に示す。補正により、ロジスティック曲線によくフィッティングしていることがわかる。

4. 理解度と振る舞いの関係性の分析

受講者の振る舞いと理解度が得られれば、それらの関係を解析する種々のアプローチが考えられる。現状では、2節で述べた受講者の姿勢系列推定は、精度が十分でないため、受講者映像を見ながら手作業で振る舞いをタグ付けした。振る舞いとしては、以下の5つを採用した。

能動的受講行動 ノートテイキングなど、自ら積極的に講義に参加している状態

受動的受講行動 顔上げや講義資料参照など、講師の教示を受動的に聞いている状態

散漫行動 よそ見やあくび・手いじりなど、講義に対する集中が途切れている状態

逸脱行動 居眠りや携帯電話操作・読書など、講義とは無関係な行動を取り、講義から意識が逸れている状態

PC 端末閲覧・操作 講義に持ち込んだノート PC 端末を閲覧、あるいは操作している状態

各受講者について講義のスパン毎に、それぞれの振る舞いの出現頻度を集計し、5次元ベクトルとして表現した。得られた5次元ベクトルをクラスタリングすることで、典型的な振る舞いをするグループに分類した。各グループについて、3節で述べた個人差、難易度差の補正をした客観的理解度の平均と、クラスター中心の振る舞い頻度に基づいて、それぞれのグループを分析した。

大学院向けの1講義5回分について、13個のスパンを設定して、理解度アンケートと小テストを実施した。講義は、最大受講者13人が受講しており、計73組のデータを得た。このデータをクラスタリングした結果、以下の5つの典型的なタイプが得られた。

受動的タイプ 受動的受講行動の頻度が非常に高く、客観的理解度も母平均よりも高い。

PCタイプ PC 端末閲覧・操作の頻度が非常に高く、客観的理解度は、受動的タイプには劣るものの母平均よりも高い。

逸脱タイプ 逸脱行動の頻度が非常に高く、客観的理解度は母平均を下回っている。

能動的タイプ 能動的受講行動の頻度が高く、客観的理解度は最も高い。

散漫タイプ 散漫行動の頻度が他のクラスターと比較して高く、受動的受講行動の頻度がやや低い。受動的タイプと比較して客観的理解度がやや低くなっているが、母平均とは等しい値である。

以上の5つのタイプから、能動的受講行動や受動的受講行動、PC操作といった行動が客観的理解度と正の相関を持ち、逸脱行動が負の相関を持つという傾向が確認できた。

5. おわりに

本稿では、講義における受講者の振る舞いと理解度との関係性を明らかにすることを目指して行っている研究について述べた。現状は、いくつかの要素技術を開発している段階であるが、これらを統合し、実際の講義において多数の受講者に適用することで、振る舞いと理解度との関係性を実証していくことが今後の目標である。

謝辞

本研究は、科学研究費 22500919 および 23300311 の助成を受けて実施した。

参考文献

- [1] “授業観察事始め 授業というフィールドにおける本格的な行動研究を目指して,” 京都大学高等教育教授システム開発センター編, 玉川大学出版部, pp.74-99 (2001).
- [2] “授業課程の評価指標としての学生の「顔上げ」行動,” 京都大学高等教育教授システム開発センター編, 玉川大学出版部, pp.99-119 (2001).
- [3] “Kinect- XBox.com,” <http://www.xbox.com/kinect>.
- [4] Ferrari, V. Marin-Jimenez, M., Zisserman, A.: “Progressive Search Space Reduction for Human Pose Estimation,” CVPR 2008, pp.1-8 (2008).
- [5] 大友賢二: “項目応答理論 TOEFL・TOEIC等の仕組み”, 電子情報通信学会誌, Vol. 92, pp. 1008-1012 (2009).
- [6] 小山昂祐, 廣瀬達, 大野公裕, 原田史子, 島川博光: “項目反応理論を用いたプログラミング演習における理解度確認”, 第9回情報科学技術フォーラム(FIT)論文集, Vol. 3, pp. 623-624 (2010).