

心的状態の持続モデルを考慮した生体情報からの学習者の 心的状態推定の試み

Trial Study on Estimation of Learner's Mental States from Physiological Indexes Considering Persistent Model of Mental States

松居辰則¹ 宇野達朗² 田和辻可昌¹

Tatsunori Matsui¹, Tatsuro Uno², and Yoshimasa Tawatsuji¹

¹ 早稲田大学人間科学学術院

¹ Faculty of Human Sciences, Waseda University

² 早稲田大学人間科学部

² School of Human Sciences, Waseda University

Abstract: Estimation of learners' mental states during the interaction between teacher and learners is very important issues for teacher from quality of learning environment point of view. In this experimental study, relationship between teacher's utterances, behaviors, learner's physiological indexes and mental states were tried to be detected by Machine Learning Method. Particularly, in this study, we tried on estimation of learner's mental states from physiological indexes considering persistent model of mental states. As a results, the effectiveness of considering persistent model and time delay of physiological indexes were suggested.

1. はじめに

教授・学習過程において学習者の心的状態を把握することは教育効果・学習効果の観点から極めて重要である。教育工学研究においても、学習者の眼球運動や発汗量など生体情報を学習行為や心理状態と関係付けるための基礎的な研究は多くの知見を蓄積している。そして、昨今の計算機や生体計測機器の高機能化と低廉化によって、生体計測機から得られるリアルタイムかつ大量のデータを高速に処理することにより、生体情報や行動情報を用いた学習者の心的状態の計算機による自動推定と教育支援への試みが盛んに行われている。一方、教育実践研究においては、教授・学習過程における教師と学習者のインタラクションは学習者の心的状態の変化に影響を及ぼし、学習効果の決定要因として重要であることは広く共有されているところである。したがって、学習時における教師の行動や発言と学習者の心的状態、あるいは心的状態の変化に関する要因との関係の形式化は重要な課題であり、その成果は学習支援システムへの学習者の心的状態の推定機能の実装のための基礎的な知見を与えることも期待される。

著者らは、教師と学習者のインタラクションにおいて教師の発話と学習者の生理データ、および学習

者の心的状態との関係の形式化を試みてきている[1]。そこでは、生体情報、発話情報、自省報告等、学習に関わる情報を多面的に用い相関ルールにより関係の形式化を試みている。一方、機械学習アルゴリズムの学習支援を含めて様々な分野への応用可能性も示されてきている[2][3][4]。そこで、本研究では、上記と同様の学習に関わる多面的情報から Deep Neural Network (以下、DNN) を用いて学習者の心的状態の推定の可能性を検討した。本研究は、著者らが開発を進めている知的メンタリングシステム (IMS: Intelligent Mentoring System) [2]の基盤となる研究と位置付けられる。

2. 学習に関わる多面的情報の取得

学習に関わる多面的情報の取得を目的として生体計測機器を用いた計測を中心に実環境での実験を行った。被験者は個別指導塾(教師1名、学習者1名の個別学習)に通う中学生1名であった(本実験の実施にあたっては当該塾の講師を通して被験者(生徒)および保護者の許可を得ている)。使用した機材は Near-Infrared Spectroscopy (NIRS) (日立 WOT-100製)、呼吸・皮膚コンダクタンス・容積脈波計 (NeXus) であった。被験者には上記の機材を全て装着しても

らい、通常通りの授業を受けてもらった。各計測機器は計測時間を対応させるために計測開始、終了時にマーカーを付与した。得られたデータのうち、NIRS のデータに対して、大域平均基準化[5][6]を施した。実験中の様子は3か所から3台のビデオカメラで撮影した。この映像から、著者らが教師の教授行動、特に発話行為に焦点を当て、「説明」「指示・確認」など9種類のカテゴリに分類した。教師の発話を表すカテゴリは、先行研究[7][8][9]で使用されていたカテゴリを一部修正した9種類のカテゴリ（1：説明，2：発問，3：指示確認，4：復唱，5：感情受容，6：応答，7：注意，8：雑談，9：その他）を用いた。教師の発話へのカテゴリの付与に関しては、分析者が授業映像を見ながら分析者の視点で行った。また、被験者には後日実験で得られた映像を見ながら学習時の心的状態の内省報告を求めた。心的状態を表すカテリは、Achievement Emotions Questionnaire (AEQ) [10]で使用されている9感情についての尺度 (Enjoy, Hope, Pride, Anger, Anxiety, Shame, Hopelessness, Boredom, Other) を用いた。被験者にはアノテーション付与のための自作アプリケーション（自身の動画を再生しながら該当する心的状態のボタンを押すことによってその時の心理状態を内省報告する）を用いて授業時の心的状態の内省報告を求めた。

3. 3 層ニューラルネットワークを用いた心的状態推定の試み

2 節で取得したデータを用いて 3 層ニューラルネットワークを用いた生体情報からの心的状態の推定を試みた。今回分析の対象としたのは、約 60 分授業の中で教師と学習者のインタラクションが比較的多く確認できた 63 秒（開始後 19 分 37 秒から 20 分 40 秒まで）であった。

3.1 データ構造とネットワーク構造

入力データは、①NIRS によって取得した脳血流量 (5Hz)、②呼吸 (32Hz)、③皮膚コンダクタンス (32Hz)、④教師の発話 (上記の 9 カテゴリのうち「説明」「発問」「指示・確認」「注意」「その他」) であった。出力データは、⑤内省報告 (上記の 9 カテゴリ) であった。なお、①②③は粒度を統一するために、粒度の荒いデータに対して、粒度の最も細かいデータの粒度に合わせて線形補完を施した。その結果、最終的な対象データ数は、2024 行 × 5 (「NIRS」「皮膚コンダクタンス」「呼吸」「教師の発話」「心的状態」) となった。また、入力データのうち、①②③の数値データは、それぞれカテゴリ内で標準化 (平均 0、

分散 1) をおこなった。

ネットワーク構造は、入力層・中間層 (1 層)・出力層の 3 層ニューラルネットワークとした。中間層のユニット数はシミュレーションの結果 19 とした。これは、中間層のユニット数を 1 から 25 まで変化させたところ、ユニット数 19 において、学習の Loss 値が最小、Accuracy 値が最大となったためである。この結果を図 1 に示す。

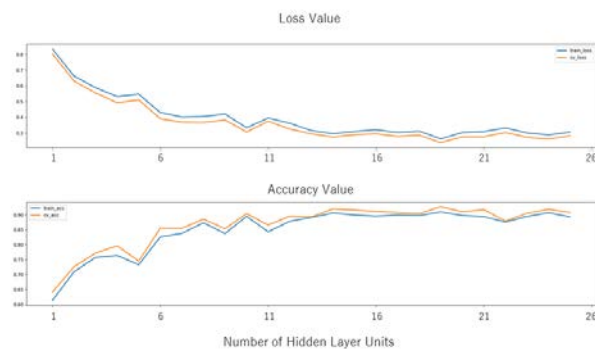


図 1 中間層ユニット数による Loss 値, Accuracy 値の変化

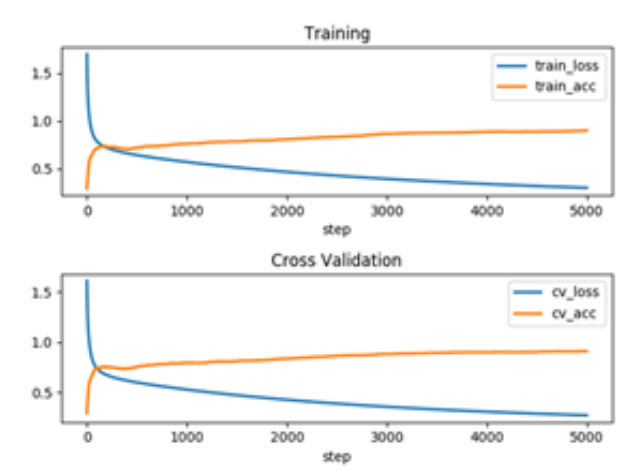


図 2 3 層 NN におけるシミュレーション結果

3.2 シミュレーション方法と結果

Python3.5, Tensorflow (ver 0.12.1) で実装した。活性化関数は中間層では tanh 関数を、出力層ではソフトマックス関数を用いた。また、損失関数にはクロスエントロピー誤差関数、オブティマイザには Gradient Descent (最急降下法) を用いた。学習率は 0.05 とした。対象データ (2049 データ) を 6:4 で分割し、6 割のデータを学習データ、4 割のデータを評価データとして 10 回の交差検定 (Cross-Validation) を行った。また、学習にあたっては 5000 回繰り返し学習させた。学習では交差エントロピーが最小になるように学習させた。その一部を図 2 に示す。早い段階で学習が収束していることがわかる。また、10

回の交差検定において Accuracy 値を算出(小数点第4位で四捨五入)した結果、「0.900, 0.917, 0.910, 0.917, 0.910, 0.914, 0.915, 0.899, 0.912, 0.905」となった。このことから、非常に高い精度で生体情報から心的状態の推定が行われていることがわかる。

4. DNN を用いた心的状態推定の試み

3 節で説明した三層ニューラルネットワークでは、時間的側面が考慮されておらず、ある一つの単位時刻における生体情報と心的状態の関連性について議論されてきた。そこで、この点を考慮したうえで、中間層を多層化することで推定の精度向上が認められるかを検討した。

4.1 リアルタイム性処理と時間的側面の検討

前節では、ある一つの単位時刻における生体情報と心的状態に関する入出力の写像関係を三層のニューラルネットワークを用いて学習させた。ところが、このニューラルネットワークでは、以下の二点において課題が残る。一点目は、応用的側面を考慮すると入力データの標準化を行なうことが難しいという点である。より具体的には、IMS などのリアルタイム性の高い処理が求められる状況では、NIRS などから得られた情報に対して大域平均基準化や、生体情報に対する標準化を適用することが難しい。したがって、IMS を検討する上では、リアルタイムに入力される NIRS などの生体情報データをそのまま使用する必要がある。二点目は、生体情報の表出および教師の発話があった時刻から心的状態の認知に至るまでの時刻に存在すると考えられる時間遅れが検討されていない点である。そこで、入力から出力に至るまでの時間遅れを考慮した入出力の関係を学習する必要がある。

4.2 データ構造とネットワーク構造

入力データは3節と同様、インタラクションが多く確認できた63秒間に存在する①NIRS、②呼吸(32Hz)、③皮膚コンダクタンス(32Hz)、④教師の発話であった。ただし、DLでの学習においては、生体情報に関するデータの標準化は行われていない。また、出力データは3節と同様に⑤内省報告であった。ネットワーク構造は、入力層・中間層(2層)・出力層の4層ニューラルネットワークとした。時間遅れを表現するために、入力層では時刻 $t-\tau$ から時刻 t までの①②③④のデータを、また、出力層では時刻 $t-\tau$ の心的状態を出力することで時間遅れを表現した(図3)。尚、中間層のユニット数は中間層1と

中間層2のユニット数を変化させながらシミュレーションを行い、Loss 値と Accuracy 値を比較することで、一層目19次元、二層目を20次元と決定した。図4に、中間層2のユニット数を20として、中間層1のユニット数を段階的に変化(10~19)させながら、シミュレーションを行った際の Loss 値と Accuracy 値を示す。また、本シミュレーションでは、3および10ステップ、つまり、 $\tau=2,9$ でシミュレーションを行ったが、 $\tau=2$ のときは、中間層1のユニット数を変化させても、Accuracy 値が、 $\tau=9$ のときを超えることがなかった。したがって、以降本シミュレーションでは、 $\tau=9$ の場合について述べる。

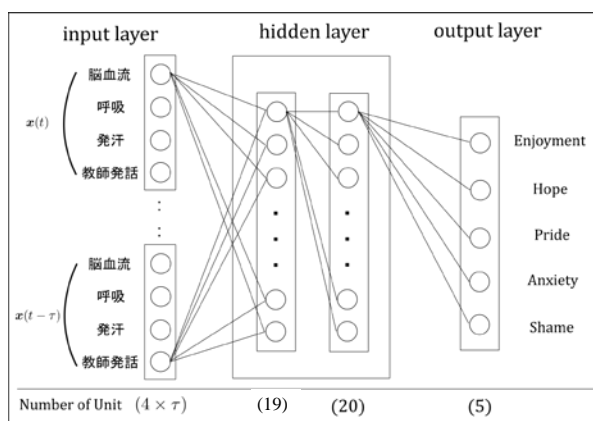


図3 ネットワーク構造図

4.2 シミュレーション方法と結果

Python3.5, Tensorflow(ver 0.12.1)で実装した。中間層の活性化関数は二層とも tanh 関数を、出力層ではソフトマックス関数を用いた。また、損失関数にはクロスエントロピー誤差関数、オプティマイザには Gradient Descent(最急降下法)を用いた。学習率は0.05とした。

中間層1のユニット数19, 中間層2のユニット数20でシミュレーションを行った。対象データ(2049データ)を6:4で分割し、6割のデータを学習データ、4割のデータを評価データとして10回の交差検定(Cross-Validation)を行った。また、学習にあたっては5000回繰り返し学習させた。学習では交差エントロピーが最小になるように学習させた。その一部を図5に示す。早い段階で学習が収束していることがわかる。また、10回の交差検定において Accuracy 値を算出(小数点第4位で四捨五入)した結果、「0.799, 0.842, 0.797, 0.793, 0.764, 0.811, 0.789, 0.829, 0.768, 0.808」となり、平均0.800, 標準偏差0.023となった。このことから、入力データを標準化していなくとも0.800程度の精度で生体情報から心的状態の推定が可能であることが示唆された。

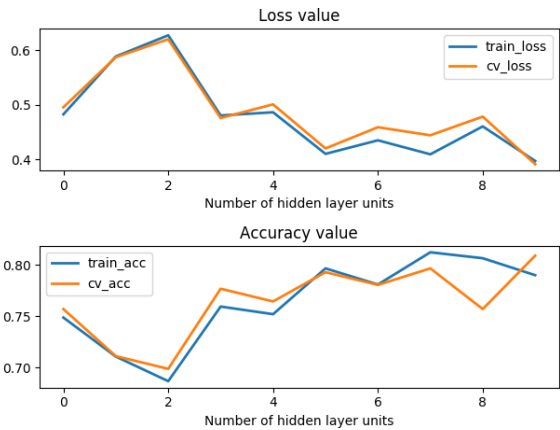


図4 中間層1のユニット数を10から19まで段階的に変化したときの Loss 値と Accuracy 値

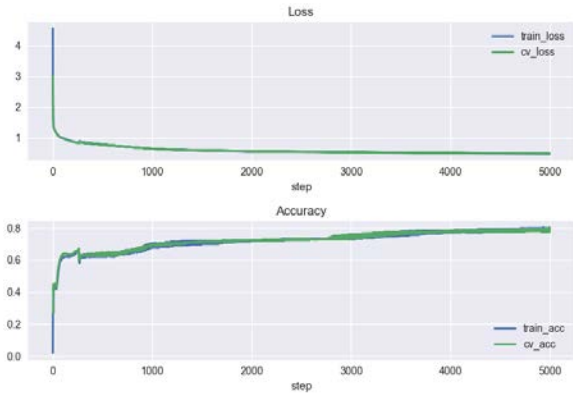


図5 5000回学習時の Loss 値と Accuracy 値の変化

5. 次元解析による時間遅れと次元の決定

これまでの、時間遅れ τ (生体情報の表出および教師の発話があった時刻から心的状態の認知に至るまでの時刻に存在すると考えられる) については、経験的に決定していた ($\tau=3\sim 10$)。そこで、次元解析の手法[11][12] (図6参照) を用いて、時間遅れ τ とその自由度 (次元) を求めた。具体的には、各生体情報を時系列データとして捉え、その自己相関性が最も高くなる場合の時間遅れ τ と、システムとしてもつ次元 (システムの自由度: そのシステムをモデル化の際に必要な最低限の局所的な独立変数の数度) を求めた。その結果を表1に示す。

この結果から、①NIRSは2次元、②呼吸は1次元、③皮膚コンダクタンスは4次元、④教師の発話をDLの入力データとした (合計1530データ)。出力データは⑤内省報告であった。なお、時間遅れを考慮した結果、推定される感情状態 (内省報告) は3種類となった (これまでの5種類)。

ネットワークの構造は4節と同様であり、中間層

1のユニット数は19、中間層2のユニット数は20であった。中間層の活性化関数は二層とも tanh 関数を、出力層ではソフトマックス関数を用いた。また、損失関数にはクロスエントロピー誤差関数、最適化には Gradient Descent (最急降下法) を用いた。学習率は0.05とした。対象データ (1530データ) を6:4で分割し、6割のデータを学習データ、4割のデータを評価データとして10回の交差検定 (Cross-Validation) を行った。また、学習にあたっては5000回繰り返し学習させた。学習では交差エントロピーが最小になるように学習させた。その一部を図7 (上段 Loss 値, 下段 Accuracy 値) に示す。今回も4節の場合と同様に早い段階で学習が収束していることがわかる。また、10回の交差検定において Accuracy 値を算出 (小数点第4位で四捨五入) した結果、平均0.894、標準偏差.0087となった。このことから、次元解析による時間遅れとシステムの次元を考慮した結果、4節の場合と比較してより高い精度精度で生体情報から心的状態の推定が可能であることが示唆された。

表1 次元解析の結果 (時間遅れ τ と次元 m)

	NIRS	呼吸	皮膚コンダクタンス
τ (Step)	400	155	173
次元 m	1.86	1.18	3.81

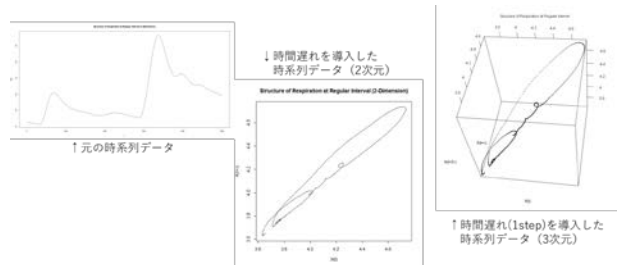


図6 次元解析のイメージ

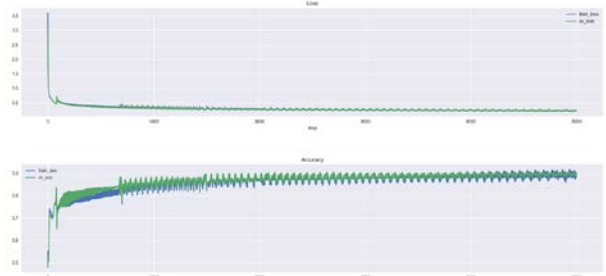


図7 5000回学習時の Loss 値と Accuracy 値の変化 (次元解析による時間遅れとシステムの次元を考慮)

6. 心的状態の持続モデルの導入

一般に人間の心的状態は時間経過に従って変化するとされている。特に、ある刺激に対して生じた感情（心的状態）は、時間の経過に対して指数関数的に減衰するとの知見が報告されている[13][14].

6.1 心的状態の持続モデル（感情強度関数）

そこで、本研究においては、④教師による発話と⑤学習者の心的状態に対して、以下のような減衰関数を適用することにした。

$$f(t) = f(T_0) \times \exp\left(-\frac{t-T_0}{T_1-T_0}\right) \quad (T_0 \leq t \leq T_1)$$

ただし、 T_0 、 T_1 はそれぞれ、教師のある発言、および学習者のある心的状態の開始時刻と終了時刻を表す。 $f(t)$ は emotional intensity（感情強度）を表現する関数であると考えられる。なお、本研究においてはデータの性質上、 $f(T_0)=1$ としている。

6.2 3層 NN, DNN の感情強度関数の適用

5 節までに用いたデータに、感情強度関数を適用して、3層 NN, DNN による、生体情報と教師の発話からの学習者の心的状態の推定を試みた。5 節までと同様に、入力データは①NIRS, ②呼吸, ③皮膚コンダクタンス, ④教師の発話であった。また、出力データも同様に⑤学習者の心的状態であった。

ネットワークの構造は4節と同様であり、中間層1のユニット数は19, 中間層2のユニット数は20であった。中間層の活性化関数は二層とも tanh 関数を、出力層では出力ベクトルの max 値以外を0として最小二乗法を用いた。また、損失関数にはクロスエントロピー誤差関数、オプティマイザには Gradient Descent（最急降下法）を用いた。学習率は0.05とした。対象データ（1530データ）を6:4で分割し、6割のデータを学習データ、4割のデータを評価データとして10回の交差検定（Cross-Validation）を行った。また、学習にあたっては5000回繰り返し学習させた。学習では交差エントロピーが最小になるように学習させた。

時間遅れを考慮したモデル（3層 NN, 中間ユニット数12）の Loss 値の変化と時間遅れを考慮したモデル（DNN, 中間層層数:3（ユニット数(20,3,12)）の Loss 値の変化を表2, 表3に示す。

表2 5000回学習後の Loss 値（3層 NN）

Alpha	最適化手法	Gradient Descent	Adam
0.05		0.0526	0.0526
0.01		0.0526	0.0526
0.005		0.0526	0.0526
0.001		0.0526	0.0526
0.0005		0.0526	0.0526
0.0001		0.0555	0.0549

表3 5000回学習後の Loss 値（DNN）

Alpha	最適化手法	Gradient Descent	Adam
0.05		0.0526	0.0526
0.01		0.0526	0.0526
0.005		0.0526	0.0526
0.001		0.0526	0.0526
0.0005		0.0526	0.0526
0.0001		0.0557	0.0526

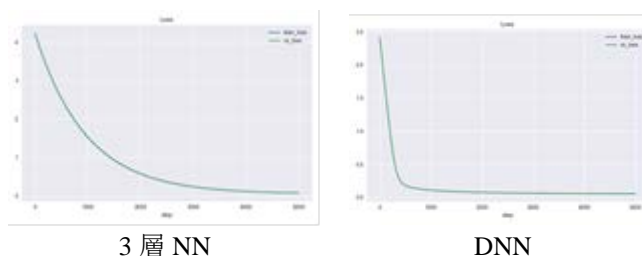


図8 Loss 値の変化（学習係数:0.0001/オプティマイザ: Gradient Descent）

いずれの場合も、Accuracy 値は0.950程度の（Loss 値は0.050程度）高い値に収束している。この結果は、心低状態の持続モデルを考慮しない場合の値（0.890程度）と比較しても高い値になっている。

さらに、図8に3層 NN と DNN との場合の Loss 値の変化の様子を示す（学習係数:0.0001/オプティマイザ: Gradient Descent）。この結果、DNNの方が3層 NNに比較して、Loss 値の収束が顕著に早いことがわかる。この結果は、オプティマイザ（例えば Adam 等）によらず同様の結果となった。また、学習係数の値の増加に伴って、この傾向はより顕著になるとの結果も得た。

7. 考察

DNNを用いた本シミュレーションから、標準化していない生体情報から約0.800の心的状態が推定で可能であることが実験的に確認された。今回、生体情報における時間遅れの程度を次元解析の手法で求め、DNNへの入力データに反映させた。その結果、より高いAccuracy値での心的状態の推定が可能となった。さらに、心的状態の持続に関するモデルを反

映させた結果, さらに高い Accuracy 値で心的状態の推定が可能になった. つまり, 計測データをより実現象を近似するデータに加工することの有効性を知見として得ることができた.

8. まとめと今後の課題

本研究では, 学習に関わる多面的情報から学習者の心的状態の推定を試みた. 具体的には, DNN を用いて学習時の生体情報と心的状態の関係のシミュレーションを行った. 特に, 次元解析による生体情報の時間遅れと自由度 (システムの次元), 心的状態に持続性に関するモデル (感情強度関数) を導入することにより, より高い Accuracy 値で心的状態の推定が可能となるとの知見を得た.

一方で, 本シミュレーションでは, 学習データにおける感情状態の少なさから, 過学習が起こっていることが示唆された. したがって, 今後は今回の 63 秒のデータに加え, 実験によって取得した全体のデータを使用することも検討する必要があると考えられる. また, 今回の DNN による学習では中間層の層数とユニット数が深く吟味されていない. 今後は様々な中間層のパターンを検討した上でのネットワーク構築が望まれる. さらに, 今回の結果から機械学習による心的状態の推定の可能性が示唆されたが, 学習・教育文脈でのデータの解釈には因果関係に関する説明が必要である. この観点からは, DNN の中間層や重みの状態の可視化と解釈が重要な課題である. 著者らは先行研究において, 学習に関わる多面的情報, つまり, 教師の発話, 学習者の心的状態, 学習者の生体情報との関係性の形式化を相関ルール抽出の手法を用いて試みている. 今後は, これらの知見を融合させることで DNN の中間層の解釈の可能性が高まるものと考えている.

謝辞

本研究の一部は, 基盤研究 (C) (科学研究費補助金)・課題番号 (16K00367), 科学技術振興機構 (JST) 事業・問題解決型サービス科学研究開発プログラム・「高等教育を対象とした提供者のコンピテンシーと受給者のリテラシーの向上による共創的価値の実現方法の開発」によって行われた.

参考文献

[1] Takehana, K. & Matsui, T. (2016. July). Association Rules on Relationships Between Learner's Physiological

Information and Mental States During Learning Process, In Proceedings of HCI International (HCII2016), LNCS 9735, 209-219.

- [2] 藤好宏樹, 吉村和代, K. Kunze & 黄瀬浩一. (2015). 英文問題解答時の視点情報を用いた英語能力推定法. 信学技報, 115(24), 49-54.
- [3] 堀口祐樹, 小島一晃 & 松居辰則. (2010, March). MRA を用いた学習者の Low-Level Interaction 特徴からの行き詰まりの推定手法, 第 58 回人工知能学会先進的学習科学と工学研究会資料, A903, 1-6.
- [4] 小島一晃, 村松慶一 & 松居辰則. (2014). 多肢選択問題の回答における視線の選択肢走査の実験的検討, 教育システム情報学会誌, 31(2), 197-202.
- [5] 野澤孝之 & 近藤敏之. (2009. September). NIRS 脳計測データのオンライン分析のためのアーティファクト除去手法の比較, 計測自動制御学会生体・生理工学シンポジウム論文集, 24, 381-384.
- [6] 平山健太, 綿貫啓一, 楓和憲. (2012). NIRS を用いた随意運動および他動運動の脳賦活分析, 日本機械学会論文集 C 編, 78(795), 3803-3811.
- [7] 藤江康彦. (2000). 一斉授業における教師の「復唱」の機能: 小学 5 年の社会科授業における教室談話の分析, 日本教育工学雑誌, 23(4), 201-212.
- [8] 岸俊行 & 野嶋栄一郎. (2006). 小学校国語科授業における教師発話・児童発話に基づく授業実践の構造分析, 教育心理学研究, 54(3), 322-333.
- [9] 清水由紀 & 内田伸子. (2001). 子どもは教育のディスコースにどのように適応するか: 小学 1 年生の朝の会における教師と児童の発話の量的・質的分析より, 教育心理学研究, 49(3), 314-325.
- [10] Pekrun, R., Goetz, Frenzel, A. C., Barchfeld, P. & Perry, R. P. (2011). Measuring Emotions in Students' Learning and Performance: The Achievement Emotions Questionnaire (AEQ), Contemporary Educational Psychology, 36(1), 36-48.
- [11] 鰐岡直人: 生体信号解析, http://ww7.enjoy.ne.jp/~nburika/time_series.html (最終閲覧日: 2017/06/18)
- [12] 石橋秋彦, 合原一幸, 小谷誠: 脳・神経系のカオスとフラクタル次元解析, 医用電子と生体工学, 26(1), 57-61 (1988)
- [13] Steephan, J.(2013). HED: A computational model of affective adaptation and emotion dynamics, IEEE transactions on affective computing, 4(2), pp.197-210.
- [14] 金子迪大, 堀毛一也. (2016). ポジティブ情動喚起刺激の不確実性と重要性が情動持続に与える影響, 感情心理学研究, 24(1), pp.33-41.