JAIST Repository
https://dspace.jaist.ac.jp/

Title	PC利用学習時の表情由来の心理情報抽出に関する研究
Author(s)	平子，温
Citation	
Issue Date	2019-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15930
Rights	
Description	Supervisor:長谷川　忍，先端科学技術研究科，修士（情報科学）

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

修士論文

PC 利用学習時の表情由来の心理情報抽出に関する研究

1610427　平子　温

主指導教員　　長谷川　忍

審査委員主査　長谷川　忍

審査委員　　白井　清昭
小谷　一孔
岡田　将吾

北陸先端科学技術大学院大学
先端科学技術研究科
（情報科学）

平成 31 年 2 月

概　要

In recently, development of ICT is changing our daily life. For instance, applying ICT to an educational field is very active in not only personal learning but also school management. Especially, e-learning (applying internet technology to education) is a famous one.

One of the challenges of e-learning is about the motivation of learning. Nowadays, there exist a couple of smartphone apps for education or e-learning. Most of the apps focus on improving learner's motivation based on the forgetting curve theory, gamification technique, and/or habituation system with a notice function of smartphone. We thought that a method which predicts learners' motivation in their learning process makes other research assist about motivation management under a hypothesis that their motivation in learning has a relationship with of their facial expression.

In order to investigate this hypothesis, first, we got a pre-experiment that three staffs attended from outside research institute of psychology as experiments participants. The pre-experiments aimed to confirm how to get accurate data for the main experiment. We did an exam with the questions about shape and regularity from the CAB test and took the videos of facial expression and PC monitor of the participants in solving the questions. After solving all questions or time up, we asked the participants to answer the questionnaire which includes “difficulty,” “interest,” “fatigue,” and “concentration”. These learners' viewpoints might reflect on their motivation to each question of the exam. After finishing the pre-experiment, we picked up the time (frame number) of solving each question from the video of their PC monitor and divided the video into a set of still images as an input resource for learner's motivation prediction.

As a result of the pre-experiments, we found a couple of issues such as the definition of the question sentences was sometimes obscurity; the participants could not remember how felt they tried such questions, limitation of the PC spec made mismatch of video frames between facial expression and PC monitor, etc.

Next, we got the main experiment which joined 19 students of JAIST and staffs of the outside research institute of psychology. We fixed the procedure of the experiment based on findings of the pre-experiment but conducted it in a similar way. We analyzed all still images (around 3,000 images) using face++, and got parameters for each [emotion : sadness, natural, disgust, anger, surprise, fear, happiness], [right and left eye gaze : position of coordinate - (x, y) , vector component - (x,y,z)], [head pose : pitch, roll, yaw], [face rectangle : width, top, left, height], [mouse status: open, close, other, mask], [static data: old, gender, glass].

Next, we calculated [average, variance, max, min] for all the parameters in each question and compressed the parameters from 3,000 to 453.

We developed a neural network in which inputs were 453 parameters and outputs were the results of the questionnaire. We devided input data into 423 train-data and 30 test-data randomly. And, we compared the average accuracy of the neural networks which had 2~7 hidden layers with “sigmoid” activation function and an output layer which had “soft-max” activation function in learning only the train-data set. We repeated this procedure 50 times to change test-data and calculated the average accuracy of the prediction.

The results of prediction were [“difficulty” max average accuracy was 0.41], [“interest” max average accuracy was 0.55], [“fatigue” max average accuracy was 0.522], [“concentration” max average accuracy was 0.505].

Especially, the max average accuracy of “interest” was 2 hidden layers. This means facial expression and “interest” of the questions had a simple correlation. We can apply other prediction methods such as (RNN, CNN, K-nn, SVM, standard statistical method). We denoted them as future tasks.

In conclusion, these results were better than just random, but needed improvements from the accuracy point of view. In the future, we want to adopt an online experiment system to increase the amount of data for machine learning input.

Keywords: facial expression., motivation, education, neural network, machine learning,

目次

第 1 章 はじめに 1

第 1.1 節 学習における PC 利用の隆盛 1

第 1.2 節 課題 1

第 1.3 節 本研究の目的 2

第 2 章 関連研究 3

第 2.1 節 e-learning における学習者の動作観測に基づく主観的難易度の推定 [7] 3

第 2.2 節 e-learning における学習時の潜在的な意識変化の抽出 [8] 4

第 2.3 節 モーションセンサを用いた学習活動の状態推定手法の開発 [9] 4

第 2.4 節 E-ラーニングのためのバイオセンシング研究 [10] 5

第 3 章 予備調査 6

第 3.1 節 予備調査の目的 6

第 3.2 節 予備調査の方法 6

第 3.2.1 項 概要 6

第 3.2.2 項 使用する外部 API に関して 6

第 3.2.3 項 使用する学習課題に関して 7

第 3.3 節 予備調査の概要 9
第 3.4 節 予備調査によって発覚した課題 10
第 3.4.1 項 アンケートの基準の問題 10
第 3.4.2 項 アンケートの想起の問題 11
第 3.4.3 項 動画のフレーム加工に関する問題 11
第 3.4.4 項 入手データの拡充 13
第 4 章 本調査 14
第 4.1 節 本調査の目的 14
第 4.1.1 項 推定の方法に関して 14
第 4.2 節 本調査の概要 15
第 4.3 節 本実験の結果 17
第 5 章 実験 18
第 5.1 節 実験目的 18
第 5.2 節 データの事前加工 18
第 5.3 節 使用したライブラリに関して 20
第 5.4 節 予備実験 21
第 5.4.1 項 隠れ層数別の比較 21
第 5.4.2 項 高学習率における隠れ層数別の比較 22

第 5.4.3 項 活性化関数別の比較 24
第 5.4.4 項 質の良いデータの選抜した場合の比較 26
第 5.4.5 項 正規化方法を変更した場合の比較 29
第 5.4.6 項 予備実験の結論 31
第 5.5 節 本実験 32
第 5.5.1 項 「難しさ」の項目の学習結果 32
第 5.5.2 項 「面白さ」の項目の学習結果 33
第 5.5.3 項 「疲労」の項目の学習結果 34
第 5.5.4 項 「集中度」の項目の学習結果 34
第 5.5.5 項 総括 35
第 6 章 他の分析手法の検討 37
第 6.1 節 統計的手法による分析 37
第 6.1.1 項 相関関係 37
第 6.2 節 RNN による分析の可能性 41
第 6.3 節 CNN による分析の可能性 42
第 7 章 おわりに 44
第 7.1 節 結論 44
第 7.2 節 今後の課題 45

第 7.2.1 項 実験段階での課題 45

第 7.2.2 項 データ分析段階での課題 47

謝辞 48

参照文献 49

付録 A　収集されたデータ 53

A1 予備調査で使用したアンケート 53

A2 本調査で使用したアンケート 55

付録 B　追試験及び新たな課題 57

第 B.1 節　概要 57

第 B.2 節　ニューラルネットワークの追試験 57

第 B.2.1 項 「難しさ」の項目における追試験結果 58

第 B.2.2 項 「面白さ」の項目における追試験結果 59

第 B.2.3 項 「疲労」の項目における追試験結果 62

第 B.2.4 項 「集中度」の項目における追試験結果 64

第 B.2.5 項 SVM・KNN 法との比較 65

第 B.3 節　GRU による推定 67

第 B.3.1 項 「難しさ」の項目における試験結果 67

第 B.3.2 項 「面白さ」の項目における試験結果 68

第 B.3.3 項 「疲労」の項目における試験結果 68

第 B.3.4 項 「集中度」の項目における試験結果 68

第 B.3.5 項 総括 69

第 B.4 節　非線形相関による分析..... 69

第 B.5 節　新たな課題 72

第 B.5.1 項 訓練データ・テストデータ分別方法の妥当性 72

第 B.5.2 項 連続値出力の方法とその評価方法の妥当性..... 73

第 B.5.3 項 RNN 及びその派生手法にかかる時間の問題..... 73

第 B.5.4 項 NN 出力層におけるクラス数問題..... 74

図目次

図 3-1 CAB 問題の例 1 （文献 [12]より抜粋） 7

図 3-2 CAB 問題の例 2 （文献 [8]より抜粋） 8

図 3-3 フレーム差の計算の概念図（FPS 一致の場合） 12

図 4-1 単純パーセプロトンの概念図 [14] 14

図 4-2 ニューラルネットワークの概念図 [14] 15

図 4-3 本調査中の PC 画面の様子（例） 16

図 5-1 データ正規化の例 20

図 5-2 学習率 0.2、シグモイド関数使用時の層数別正解率比較..... 22

図 5-3 学習率 5.0、シグモイド関数使用時の層数別正解率比較..... 23

図 5-4 学習率 10.0、シグモイド関数使用時の層数別正解率比較..... 24

図 5-5 学習率 0.2、tanh 関数使用時の層数別正解率比較..... 25

図 5-6 学習率 0.2、ReLU 関数使用時の層数別正解率比較..... 26

図 5-7 AB 判定データ使用時の層数別正解率比較..... 28

図 5-8 A 判定データ使用時の層数別正解率比較..... 29

図 5-9 正規化方法を変更したデータ使用時の層数別正解率比較 31

図 5-10 難しさの項目の層別平均正解率..... 33

図 5-11 面白さの項目の層別平均正解率 33

図 5-12 疲労の項目の層別平均正解率 34

図 5-13 集中度の項目の層別平均正解率 35

図 6-1 RNN の概念図 [14] 41

図 6-2 CNN の概念図 [14] 42

図 B-1 「難しさ」の項目追試験における平均正解率 58

図 B-2 「難しさ」項目 4 層追試験時の正解・不正解の内訳 59

図 B-3 「面白さ」の項目追試験における平均正解率 59

図 B-4 「面白さ」項目 1 層追試験時の正解・不正解の内訳 60

図 B-5 「疲労」の項目追試験における平均正解率 62

図 B-6 「疲労」項目 1 層・4 層追試験時の正解・不正解の内訳 63

図 B-7 「集中度」の項目追試験における平均正解率 64

図 B-8 「集中度」項目 1 層追試験時の正解・不正解の内訳 65

図 B-9 k 近傍法を用いた推定の平均正解率 66

表目次

表 4-1 入手したデータの一例 17

表 5-1 アンケート結果の質の判定 27

表 6-1 表情データとアンケートデータの相関係数表 37

表 6-2 表 6-1 における相関係数の集計 40

表 B-1 「面白さ」項目追試験 1 層 5000 エポックにおける予測・正答の組み合わせ 61

表 B-2 「疲労」項目の本実験と追試験における最高平均正解率の比較 . 62

表 B-3 各手法の平均正解率比較 66

表 B-4 表情の各パラメータとアンケート結果に対する MIC 69

第1章 はじめに

第1.1節 学習における PC 利用の隆盛

IT 技術、特にインターネットが 1990 年代に大きく発達し、コンピュータの所有・利用が個人のレベルまで大きく引き下げられた結果、人間のあらゆる活動においてコンピュータが利用されるようになった。教育もその例外ではなく、90 年代に CAI（computer-aided instruction、コンピュータ支援教育）というコンピュータを教育に応用するという概念が出現し [1]、00 年代にはこれにインターネットの応用を重視した e-learning（e ラーニング）という概念が主流となった [2]。

現状、e-learning という枠組みに限定せずとも、スマートフォンのアプリで英単語を学習するなど PC を教育に応用すること自体は幅広く行われている。例えば、Damien Elmes が開発した「Anki」と呼ばれるプラットフォーム [3]では外国語の例文や数学の公式などをリストで管理し、忘却曲線理論を用いて適切な（忘れそうなタイミング）で出題することで記憶の定着を支援している。このプラットフォームでは Anki のメインサーバー上に学習用の問題と学習履歴を保管し、各スマートフォン・PC の OS に対応したクライアントソフトが学習を実行することで、端末の種類を問わず、効率的な学習を提供している。

他に挙げられる試みとしては上記の忘却曲線に加えて、テストの結果を他のユーザと比較したり学習が進むにつれてユーザごとに設定されたアバターの変更可能な要素が増えたりするなどゲーミフィケーションな要素を兼ね備えた「Duolingo」 [4]や、ユーザの興味のあるジャンルに関するニュースで配信することでモチベーションを維持しつつ、ユーザがそのニュースを読んでいるときに使用可能な辞書機能の利用履歴から知らない単語・熟語を自動記録してテストを生成する「POLYGLOTS」 [5]等がある。

第1.2節 課題

これらの PC を学習に応用する試みの利点として「非同期（教師と生徒が同一の場所で同一の時間に教育活動を行う必要がない）な教育」を効率的に行うことができる事が挙げられる。e-learning 出現以前にも通信教育・遠隔教育等の概念

が存在していたが、IT 技術により成績管理の効率化や教育内容の個別化が可能な点で差がある。

しかし、欠点として「やる気が無いと続かない」「生徒側からの感情情報が少ない」[6]が挙げられる。前者は、普通の人間は学校教育を義務的に（受け身で）受けるため、自学自習に慣れていないこと、後者は PC の都合上、教師側が得られる情報の大半は文字情報であり感情の乗った情報ではないとされる。

また、私見として、前者は学校教育のように強制的な習慣で無いこと及び生徒が学習内容の重要性を理解していないケースが多いように思われる。後者においては画像・動画や音声などのデータをやり取りすることで感情情報を得ること自体は可能であるが、元々の e-learning の目的が教育の効率化であるため、画像・動画や音声などのデータを教員が解釈するための労力と時間がかかること、及び解釈自体が専門技能となることから、本末転倒になりかねないと思われる。

第1.3節 本研究の目的

本研究の目的は「学習中の生徒の表情のデータから学習に関する心理的な情報、所謂モチベーションのパラメータを推定することが出来るかどうかの検証を行うこと」である。

この目的により、第 1.1.2 項における「やる気が無いと続かない」「生徒側からの感情情報が少ない」という課題に対して、生徒側の心理情報（感情情報を含む）を、表情という生徒側の能動的な協力を要しない（負担の軽い）方法によって効率的に収集する手段を提供することで、やる気を改善する他の試みに貢献することができると考えられる。

第2章 関連研究

本章では過去に行われた本研究との類似事例を比較することにより、本研究の性質を概観する。

第2.1節 e-learning における学習者の動作観測に基づく主観的難易度の推定 [7]

この研究は一般的な学習（対面での学習）では教師は学習者の理解状況だけでなく、教材に対する関心・意欲も把握できるのに対し e-learning では対面でないため理解度は考慮されても関心・意欲が考慮されていないことを課題とし、学習者の表情・目線・頭部の向き・マウスの動きなどから学習者が教材に対してどの程度難しいと感じているか（当該論文では主観的難易度と定義している）を推定すること試み、高い精度で推定できることを示した。

本研究との関連性として、推定の材料に「表情・目線・頭部の向き」を用いる点で一致している。ただしこの 3 材料（表情・目線・頭部の向き）とは「首をかしげる・同一箇所を凝視している」といった限定的なデータをのみを採用していることや、回答までの時間という表情と関連性はないが e-learning での学習において収集しやすいデータを用いている点で相違がある。また、これらのデータから主観的難易度を推定するための技術としてサポートベクターマシンを用いており、本研究で用いるニューラルネットワーク（第 4.1.1 項にて後述）と相違がある。

また、特筆すべき点として難易度の概念を「学習者にとっての主観」として扱っている点があり、これは後述する予備実験において問題が発生した際に（第 3.4.1 項）その解決に大きく貢献した。

第2.2節 e-learning における学習時の潜在的な意識変化の抽出 [8]

この研究では非同期型の e-learning において同期型の e-learning と比べてさりげない動作や表情から学習者の理解を測る事ができないという課題に対し、マウスの動作履歴から学習者の潜在的な意識を推定することで特別な機器を用いずにリアルタイムな学習者の異常検出を目的としている。実験を重ねた結果、「難しいと感じているときにはマウスの移動速度が遅くなる」、「マウスの移動速度が早くなる場所はある程度限られている」といった知見が得られたが、学習中の学習者における「さぼり」や「行き詰まり」といった異常の判定としては、学習コンテンツごとの差によりマウスの動作履歴に大きな差が生じた結果、これがノイズとなって異常の判定を阻害した。

本研究との関連性として、学習者の潜在的な意識という抽象化された要素ではあるものの、学習のモチベーションと関連性は見られる。相違点としては推定の材料を本研究では表情由来の情報に限定するのに対し、この研究ではマウスの動作履歴を用いている。また、推定する手段として不偏分散を用いた統計的検定という単純な方法を用いている点も興味深い。

第2.3節 モーションセンサを用いた学習活動の状態推定手法の開発 [9]

この研究では、近年の MEMS(微小電気機械システム)の発達により、様々な情報機器に加速度センサを設置できるようなったという現状を鑑み、腕時計型ウェアラブル端末などに搭載される加速度センサから学習活動の状態を推定することとを目的としている。実験では学習者の右手首に 9 軸加速度センサを搭載したウェアラブル端末を取り付けた状態で「静止・キーボード操作・マウス操作・筆記・読書」といった動作を行ってもらい、その際に得られた 9 次元データを時間窓内の統計特徴量（平均・分散・主成分特徴量）を求め、3 次元データに圧縮し、得られたデータを決定木・k-NN・SVM などの分類器を用いて分類し、最後に k 分割交差検証による機械学習を行った。結果最も精度が高かった線形

SVM を分類器として選定し実験を 10 回ほど行って精度の平均値を取った結果、極めて高い精度で分類することができた。

本研究との関連性として、この研究における分類の対象として PC を用いた学習活動を想定している点がある。相違点として推定の材料に本研究では表情を用いるのに対し、この研究ではモーションセンサを用いている。特に論文前半で書かれたモーションセンサからの情報のみで装着者の姿勢や行動が推定できるという関連研究から学習に関連する活動も推定できるとした考察は興味深い。

また、PC も用いた学習活動というのに筆記が含まれているというのは、学習活動が PC 内で完結しないもの（画面を書き取るなど）も想定しており対象の範囲が広い。

他にも得られたデータの平均や分散などの統計量を取ったり、9 次元データを 3 次元データに圧縮したりするという手法は本研究においてニューラルネットワークで分類を行う際に入力データの加工を行う際(第 5.2 節にて後述)に参考にした。

第2.4節 E-ラーニングのためのバイオセンシング研究 [10]

この研究では e-learning やアクティブラーニングにおいて様々な原因による失敗（当該論文で挙げられているものでは学生側の成果物水準低下や教員側の費用負担増加等）を論じた上で、学生側の原因である目標喪失による怠惰や不挑戦、雑談や無発言等によるグループワーク無機能化などを改善のために、まずその状態を検出することを目的としている。そこで集中度（コンテンツや授業への集中の程度）と活性度（知識の応用、活用などの程度）という 2 つの基準を設け、これらが脳波や心拍、皮膚電流といった生体情報と関連していないかを検証している。現時点で研究は終了していないが、派生研究として瞬きによる筋電位の変化から集中度を測定するなどしている。

本研究との関連性として、集中度・活性度という学習のモチベーションに非常に近い概念を用いている事が挙げられる。また、この研究では推定材料として生体情報を用いているが、表情もある意味表情筋の筋電位によって生じる生体現象の一種であるため、共通点がある。

第3章 予備調査

第3.1節 予備調査の目的

本項目で説明される予備調査とは、表情と心理に関するデータを入手する調査（以下、本調査）の前に、本調査において発生しうる課題を事前に把握するために行われた調査である。その性質上、データの収集はするものの、得られたデータに対する解析は行わなかった。本項目ではその目的、手法、結果について記述する。

第3.2節 予備調査の方法

第3.2.1項 概要

予備調査はその目的上、本調査と共通であるが、第 3.1 節で述べた「表情と心理に関するデータ」と第 1.3 節で述べた「モチベーションのパラメータを推定」の観点から、表情と学習中の心理が何らかの形で連動しているという想定のもとで進める。例えば学習者が難しい課題に直面したときに首を傾げたり画面に顔を近づけたりする動作をしたり、課題に飽きてくると目線の移動が散漫になったりするなどの動きが画像上で得られると考えられる。この動きを表情分析 API で検出できないかと考えた。

したがって予備調査の進め方の概要は、まず実験協力者に学習課題を解いてもらい、その後アンケートで学習課題に対するモチベーションに関与していると思われる、学習課題に対する難しさや面白さ、回答中の疲労の程度等について回答して頂く。同時に、実験協力者が学習課題を解いている様子を Web カメラとスクリーンショットで撮影して得られた動画を外部の表情分析 API を用いて解析し、その解析結果を入力として機械学習技術を用い、アンケート結果を正解とした教師あり学習を行うものとした。

第3.2.2項 使用する外部 API に関して

第 1.2.1 項で述べた外部の表情分析 API として中国の Megvii 社が提供する「Face++」 [11]と呼ばれる API を使用した。選定理由としてこの API は現状無償で利用できることと、得られる分析の種類が他の表情分析 API と比べて多

いこと（Microsoft 製 Emotion API では表情成分の分析しかできないが、Face++では目線・口角・年齢推定なども行う）や精度が高いこと（Google 製 Cloud Vision API では各表情成分の分析値が離散 3 値で返されるのに対して、Face++は 0~100 の連続値で返される）と比べて多い事等から選定した。

第3.2.3項 使用する学習課題に関して

第 1.2.1 項で述べた習課題として、日本エス・エイチ・エル社製の就職試験問題「CAB」の「法則性」 [12]の模擬問題を使用した。この問題はある法則性に基づいて並んだ 5 つの図形の内、1 つだけ欠けている物があり、これを残り 4 つの図形から法則性を推理して合致する図形を 5 つの選択肢から選ぶという課題である。この例題を以下の図に示す。

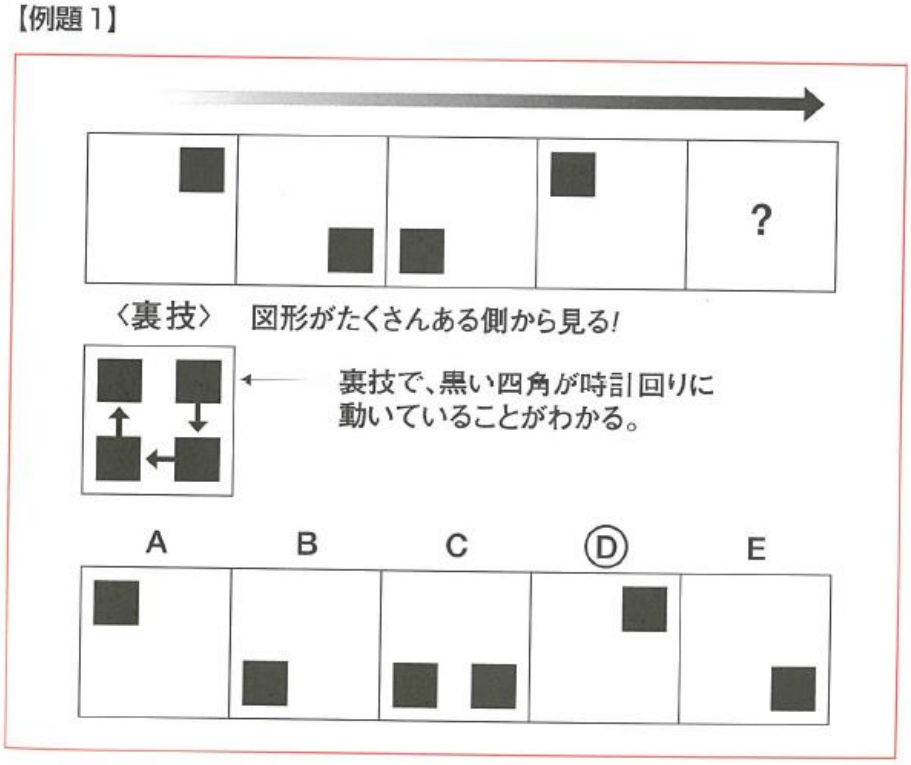

図 3-1 CAB 問題の例 1 （文献 [12]より抜粋）

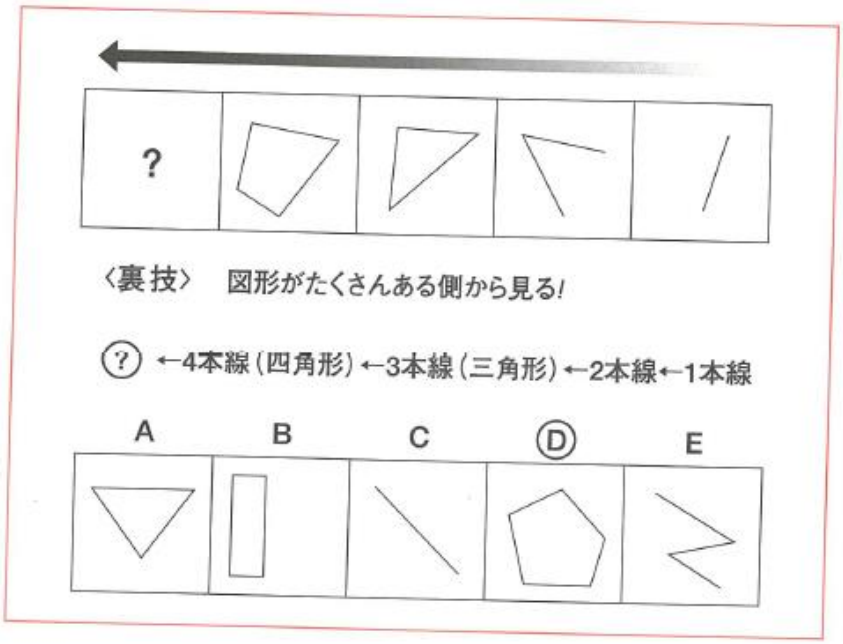

図 3-2 CAB 問題の例２（文献 [8]より抜粋）

この課題の選定理由として第１に表情から心理的な情報を時系列的に抽出するという性質上、アンケートで獲得する情報について時系列の単位が長すぎると実験協力者が正常に思い出せない可能性があり、逆に時系列の単位が短すぎれば実験協力者の負担が重すぎる可能性がある。このため、一問を解く時間が約１５秒～６０秒程度である必要があった。

CAB の法則性問題は（正規のものは）４０問を１５分で解くため、１問あたりの時間が約 22.5 秒であり、適切であると考えられた。候補に上がっていた他の CAB の問題（「暗算」「命令表」「暗号解読」）については、１問あたりの時間が極端に短かったり（最短で９秒）長かったり（最長で９０秒）で、同じく候補に上がった SPI などの模擬問題は長過ぎた（SPI は解答者ごとに問題総数が可変であるため１問あたりの時間が確定しないが、およそ９０～１８０秒ほどかかると言われる）。

第２に他の課題は１つの出題に対して問が３つ有るなど、問題の範囲が二重になっており、アンケートにおける問題の指定において混乱が生ずる可能性があった。例えば CAB の暗号解読問題はある図形を変形する系統図に対してその系統がいかなる処理を行っているかを推測し問いに回答するが、１つの系統図

に対して３つの問が存在し、アンケートにおいて１問目の難易度を質問するときに系統図の理解の難易度か問の難易度かで混乱する恐れがあった。

第３にこの実験を留学生の多い JAIST 内で行う可能性があることから問題の内容ができるだけ言語に依存しない内容である必要があった。他の候補に関しても図形処理など言語に依存しない問題があるが、問題の解説にあたって最も言語に依存しないと考えられたため選定した。

第3.3節 予備調査の概要

予備調査は 2018 年 7 月頃に外部の心理学研究機関の協力の下、予備調査協力者３名（１８～２４歳、いずれも男性）に対して、以下の手順で実験を行った。

- データ収集段階

1. 実験協力者に実験の流れを説明する。
2. インカメラ付きノート PC の前に座ってもらう。
3. PC のインカメラを用いて実験協力者の顔を写した動画（以下、顔動画）と PC の画面を写した動画（以下、PC 動画）の撮影を開始する。
4. 実験協力者に CAB 問題を解いてもらう。
   （問題は全部で 30 問、制限時間を 12 分とし、それ以降は残りの問題が有っても回答を終了する）
5. 顔動画、PC 動画の撮影を終了する。
6. 実験協力者に、上記の手順４で回答してもらった問題について、各問題そのものと各問題の正解と解説を見ながら、各問題の「難しさ」「面白さ」「疲労」について１～５の５段階で評価してもらう。
   （このとき、手順４で時間切れにより回答できなかった CAB 問題の分についてはアンケートに回答しなくても良い）
7. 実験を終了する。

- データ解析段階

1. 入手した顔動画のフレームを静止画化する。
2. 手順１で得られた静止画を Face++により「表情成分（悲しみ、無表情、嫌悪、怒り、驚き、恐怖、幸福）」「両目の目線（x,y,z 軸の数値）」「画面内の顔のサイズ」を取得する。

3. PC 動画から実験協力者が各問題を解いているフレームを計測する。(例えば実験協力者 2 番が 1 問目を解いているフレームは 122～172 フレーム目の間といったフレーム数を計測する)
4. 手順 3 で得られたフレームを元に顔動画における各問題のフレーム数を計測する。
5. 手順 4 で得られたフレーム数を元に、手順 2 で得られたデータの内、何番目から何番目までが各問題に該当するかを割り出す。
6. 手順 5 で得られたデータの集合をアンケートと紐つける。

この予備調査で用いた機材は以下のようなものであった。

- Lenovo 製 E560 ノートパソコン
  - CPU：Intel Celeron 3855U 1.60GHz
  - RAM：8 GB
  - HDD：1 TB
  - モニタ ：15.6 型フル HD
  - web カメラ ：2D カメラ搭載

また、予備調査のデータ解析段階において Face++による分析を行うため、Java を用いたクライアントソフトウェアを作成した。

なお、このとき用いたアンケートの詳細は付録 A1 にある。

第3.4節 予備調査によって発覚した課題

第 2.1.2 項における予備調査を実施した結果、本実験においてデータ収集の障害となりうるいくつかの課題が確認された。

第3.4.1項 アンケートの基準の問題

実験協力者からの質問でアンケート中の「疲労」の項目についていかなる基準で付けるべきか分からないとの指摘があった。アンケート中のすべての項目については実験協力者の主観でつけて良いこととしていたが、この質問をした協力者は実験前にアルバイトをしていたため肉体的に相当の疲労があり、実験開始前から疲労している状態であった。

本来アンケートにおける「疲労」の項目は問題をとき終わった時点で実験協力者がどの程度疲労しているかを記述してもらい、回答中の疲労のたまり方や溜

まったタイミングを表情から推測することが目的であったが、この場合はじめから疲労しているため、そういったタイミングの検出自体が不可能な状態であった。

この課題に関してはあくまで生徒の心理情報を得ることが目的であり疲労の状態を定量化することは目的ではないため「はじめから疲労が溜まっていたならばその状態をアンケートに書いても問題ない」と告知した。

ただし、本件の「疲労」に限らずアンケートにおける回答の基準の問題が他に発生することは考えられるため、本実験に際してはアンケート前に主観的な基準でアンケートに回答して良い旨を伝えることとした。

第3.4.2項 アンケートの想起の問題

実験協力者からアンケートの回答中に回答した問題について思い出せないとの指摘があった。具体的には「自分の回答した問題を覚えていないため、回答中にどの程度難しさを感じたか回答できない」「そもそも自分が何問目まで回答したか覚えていない」という内容であった。

第 3.3 節にも記載の通り、各問題における実験協力者自身の心理状態を思い出しやすくするためにアンケート記入時には問題と正解を参照させているが、これは全問が参照可能なため、後半になると自分が回答した問題かどうかを区別できないとのことであった。

本件においてはアンケート回答中に撮影した PC 動画を見せた結果、アンケートに回答することができたが、これが PC 動画を見た結果正しく記憶が思い出された結果によるものか、PC 動画を見た結果記憶が無意識に変造・捏造されたものを想起したと誤認したのかは確認できない。

本実験においては第 3.3 節に記載した問題の情報に加え、実験協力者の回答とその正誤も見せること、及びアンケート開始前の説明の時点で回答していない CAB 問題の分についてはアンケートの解答欄に消し線を引くこととした。

第3.4.3項 動画のフレーム加工に関する問題

第 3.3 節のデータ解析段階の手順 4 で PC 動画のフレームから顔動画のフレームを抜き出す作業を行っていたが、この作業中に両動画の FPS が一致しない問題が発覚した。

具体的には、データの収集段階において、PC 動画と顔動画はそれぞれ別のソフトウェアで撮影されているため両動画の撮影開始時にはズレがあるが、顔動画の撮影時には動画の右下に時計が組み込まれること、及び動画の撮影開始の順番について PC 動画を先にすることにより、顔動画撮影開始の瞬間の時計が PC 動画に写っているため、この時計を基準に両動画の撮影開始時のフレームのズレを計算した。この概要を以下の図 3-3 に示す。

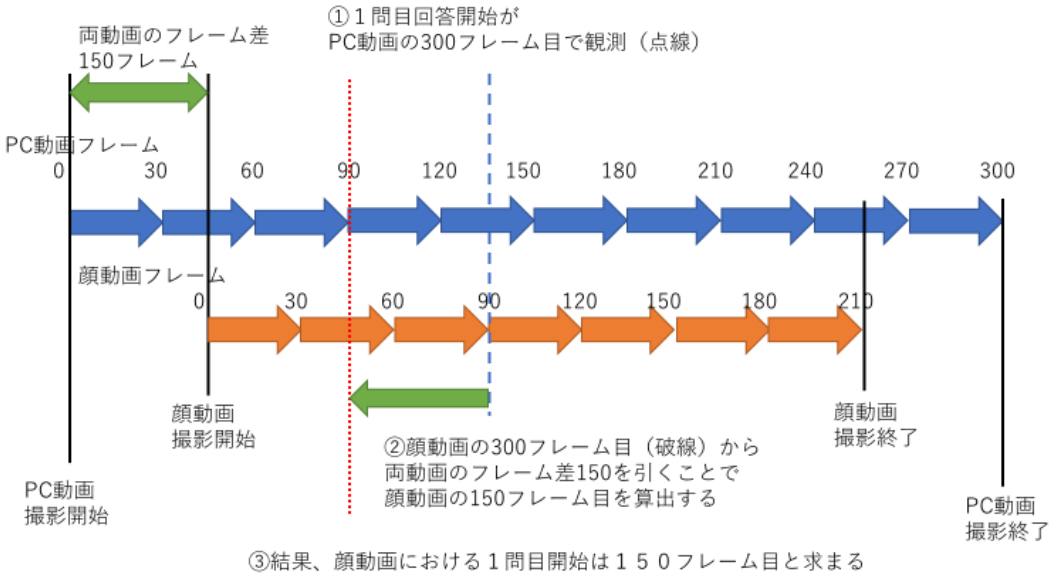

図 3-3 フレーム差の計算の概念図（FPS 一致の場合）

その後、PC 動画から各問題のフレームを抜き出した後、フレームのズレの分だけ修正を行うことで、各問題の顔動画のフレームを計算する予定だった。（例えば PC 動画と顔動画のズレが+15 フレームで、PC 動画上の 1 問目のフレームが 245~293 フレームならば、1 問目の顔動画のフレームは 230~278 と計算する予定だった）しかし上記は両動画の FPS が一致していることが前提である。実際に作業していたところ、顔動画の FPS が 29 程度なのに対し、PC 動画の FPS が 25~30 程度と低い方にばらつきがあった。このため PC 動画から計算されたフレームを顔動画に適用すると、顔動画の終端フレームよりかなり手前で最後の問題回答が終了していることが分かった。

原因としては両動画を撮影する際に、両動画の撮影ソフトウェアとも撮影設定は AVI 規格で FPS29.97 に固定されていたが、PC 動画のフレームは顔動画

のフレームよりも大きいため処理速度が低下した可能性及び PC 自体の時計の精度の悪さ等が考えられる。

このため以後の予備調査及び本調査においては、両動画の撮影開始時のフレームのズレだけでなく、撮影終了時のフレームのズレも測定することで、両動画の FPS 差を計算し、誤差を修正することとした。

第3.4.4項　入手データの拡充

実験にあたって協力した外部の心理学研究機関からの指摘で、収集するデータに以下のようなものを加えるべきとの提言を受けた。

- 静的データの収集
- アンケート項目における集中度項目の追加

前者については一般に表情というと表情筋の動きや目線の動きといった動的なデータが想定されるが、目線の動きに性差があったり、年令によって表情筋の動きに年齢差があったりするなどの問題が有り、そういた動きの差を協力者の属性ごとに区別するために静的なデータが必要との指摘を受けた。

後者については、学習におけるモチベーションを構成する要素として学習に対する集中の度合いもあるのではないかとの指摘を受けた。

この指摘に基づき、本調査の方ではアンケート項目の追加と Face++からの分析値に静的なデータも追加することとした。

第4章 本調査

第4.1節 本調査の目的

本実験は第 1.1.3 項における本研究の目的を達成するために必要な「学習中の学習者の表情のデータ」と「学習中の生徒の心理に関するデータ」及びこの組み合わせを入手することである。

第4.1.1項 推定の方法に関して

第 1.3 節で述べた「モチベーションのパラメータを推定」としてニューラルネットワークを選択した。これは人間の脳神経系の働きを数理モデルにしたもので、同様の動作を行わせることで人間と同じように問題解決能力をもたせようとするものである [13]。　具体的には複数の入力値に対して重みと呼ばれる係数を掛けた合計値を計算し、この合計値が閾値を超えたときに出力を出すパーセプトロンと呼ばれる関数をネットワーク状に組み合わせたものをニューラルネットワークと呼ぶ。この概要図を以下に示す。

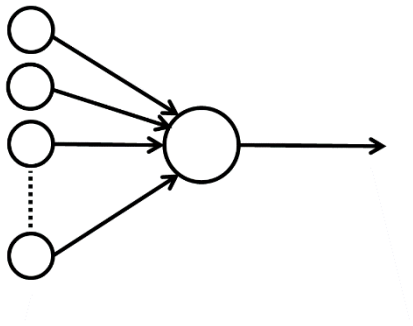

図 4-1 単純パーセプロトンの概念図 [14]

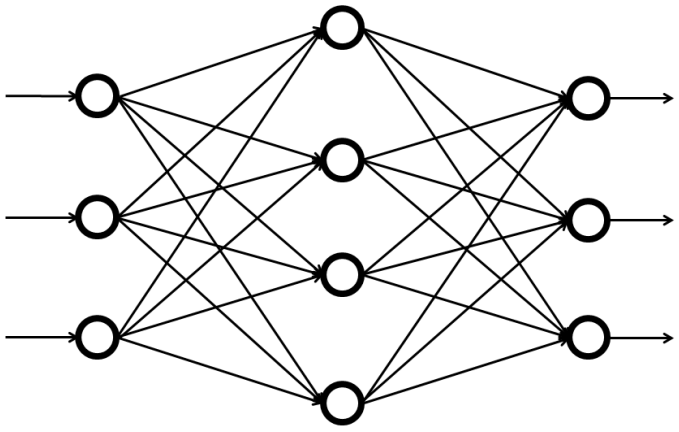

図 4-2 ニューラルネットワークの概念図 [14]

近年ではニューラルネットワークの技術を拡張した DNN（ディープニューラルネットワーク・深層学習）や CNN（畳み込みニューラルネットワーク）などの技術により、今まで分類が困難だったデータも分類ができるようになった[15]ことで注目されている。

本研究ではこれらの基本的な技術であるニューラルネットワークを主として扱う。

第4.2節 本調査の概要

基本的な手順は第 2.1.2 項における予備調査手続きと同じであるが、第 2.1.3 項に記載した予備調査において発覚した課題を元にいくつかの修正を加えた上で実施した。修正した部分を太字にした上で、本実験の手順を以下に示す。

- データ収集段階

1. 実験協力者に実験の流れを説明する。
2. インカメラ付きノート PC の前に座ってもらう。
3. PC のインカメラを用いて実験協力者の顔を写した動画（以下、顔動画）と PC の画面を写した動画（以下、PC 動画）の撮影を開始する。
4. **CAB 問題の低難易度の例題を見せ、解いてもらう。（この間にデータ解析段階における手順 4 で使う時計を PC 動画に撮影する）**
5. 実験協力者に CAB 問題を解いてもらう。

（問題は全部で 30 問、制限時間を 12 分とし、それ以降は残りの問題が有っても回答を終了する）

6. **回答終了後、アンケートの説明を行う。（この間にデータ解析段階における手順 4 で使う時計を PC 動画に撮影する）**
7. 顔動画、PC 動画の撮影を終了する。
8. 実験協力者に、上記の手順 4 で回答してもらった問題について、各問題そのものと**実験協力者の回答と正誤**、正解と解説を見ながら、各問題の「難しさ」「面白さ」「疲労」「**集中度**」について 1 ～ 5 の 5 段階で評価してもらう。（このとき、手順 4 で時間切れにより回答できなかった CAB 問題の分についてはアンケートに回答しなくても良い）
9. 実験を終了する。

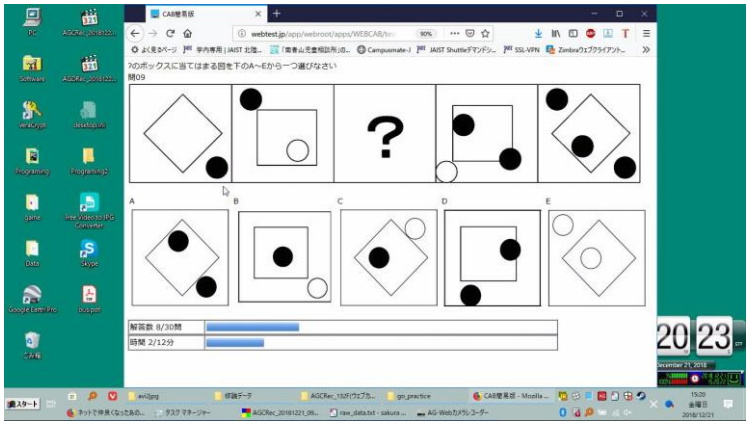

図 4-3 本調査中の PC 画面の様子（例）

- データ解析段階

1. 入手した顔動画のフレームを静止画化する。
2. 手順 1 で得られた静止画を Face++により「表情成分（悲しみ、無表情、嫌悪、怒り、驚き、恐怖、幸福）」「両目の目線（x,y,z 軸の数値）」「**頭部の向き（ピッチ・ロール・ヨー角）**」「画面内の顔のサイズ」「**口の状態**」「**推定年齢・性別・眼鏡の有無**」を取得する。
3. PC 動画から実験協力者が各問題を解いているフレームを計測する。（例えば実験協力者 2 番が 1 問目を解いているフレームは 122～172 フレーム目の間といったフレーム数を計測する）

4. PC 動画と顔動画の双方に写った時計を基準に、動画の開始時刻と終了時刻、及び両時刻におけるフレームを計測する。
5. 手順４で得られたフレームと時刻を元に FPS 差を計算する
6. 手順４と５で得られたフレームの差と FPS 差を元に顔動画における各問題のフレーム数を計測する。
7. 手順４で得られたフレーム数を元に、手順２で得られたデータの内、何番目から何番目までが各問題に該当するかを割り出す。
8. 手順５で得られたデータの集合をアンケートと紐つける。

なお、このとき用いたアンケートの詳細は付録 A2 にある。

第4.3節 本実験の結果

本実験は２０１８年６月～２０１８年１２月の間に実施した。実験協力者は外部の心理学協力機関の関係者７名と JAIST の学生１２名の合計１９名（男性１３名、女性９名、年齢２０～４６歳）が参加し、人数分のデータを得ることができた。

このとき得られたデータの概略を以下に示す。

表 4-1 入手したデータの一例

Num			Emotion				Left eye			…
P	Q	Frame	sad	anger	fear	…	Vector_x	Vector_y	…	…
2	18	787	15.003	1.664	0.328	…	0.526	0.286	…	…
2	18	788	15.556	1.432	0.427	…	0.524	0.26	…	…
3	1	312	12.118	3.839	0.328	…	0.51	0.309	…	…
3	1	313	33.75	1.895	0.517	…	0.52	0.302	…	…
3	1	314	8.454	2.481	0.315	…	0.503	0.288	…	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	…

表 4-1 の例において実験協力者が解いた問題が写っているフレームに対して協力者の番号と問題の番号によるラベル付がされた状態で、各フレームに対し、表情成分や目線などの数値が組み合わされている様子を表現している。

第5章 実験

第5.1節 実験目的

本章では第 4 章で得られたデータの内、表情に関するデータを入力、アンケートの結果を出力（正解）としてニューラルネットワークの学習を行い、生成されたネットワークを用いてアンケート結果が予想できるかどうかを確認する。

第5.2節 データの事前加工

第 3 章で得られたデータをニューラルネットワークに学習させるに当たり、入出力の数を一致させなければならない。具体的な例として、1 人目の協力者が CAB 問題の 1 問目を解いているフレームは、当該人物の顔動画の 126~341 フレーム目で、合計 215 フレームであり、このフレーム 1 つ 1 つに表情の成分や目線などのデータが割り当てられている。しかし出力は 1 つの問題に付きアンケートの出力結果が 1 つであり、データの組み合わせが入力２１５個に対して出力 1 個となっており一致しない。何らかの方法で入出力のデータ数を一致させる必要がある。

本研究ではこの 215 フレームのデータの各系列（表情の怒りの成分の数値２１５個や目線の X 方向の数値２１５個など）に対して「平均値・分散値・最低値・最高値」の 4 つの統計値を取ることで 1 個の入力データに圧縮した。

この「平均値・分散値・最低値・最高値」4 つの統計値を取った理由として、当初は平均値のみを取る予定であったが、驚きや嫌悪などの一瞬しか現れない表情だとそれが出現して驚きや嫌悪の数値が上昇したとしても一瞬であるためそれ以外の低数値によって相殺され平均値が低くなり検出できなくなる可能性があった為、最高値・最低値を追加した。

また分散については、例えば目線が左右に繰り返し動いている状態と目線が中央から全く動いていない状態は双方とも目線の平均値が中央になってしまい区別がつかないため、分散値を用いて区別することを目的に追加した。

最終的に以下のようにデータを圧縮した。

- 圧縮前　データ個数　約３０万個　３１次元
  - 表情成分　　：悲・無・嫌・怒・驚・恐・笑
  - 右目成分　　：目線Ｘ・Ｙ・Ｚ軸　黒目位置Ｘ・Ｙ軸
  - 左目成分　　：目線Ｘ・Ｙ・Ｚ軸　黒目位置Ｘ・Ｙ軸
  - 頭角度　　：ピッチ・ロール・ヨー角
  - 表情検出座標　：横幅・高さ・上からの距離・左からの距離
  - 口状態　　：閉口・開口・他・マスク
  - 静的データ　　：推定年齢・性別・メガネ
- 圧縮後　データ個数　８７次元
  - 表情成分　　：（悲無嫌怒驚恐笑）×（平均・分散・最低・最高）
  - 右目成分　　：（目線Ｘ・Ｙ・Ｚ軸）×（平均・分散・最低・最高）
  - 左目成分　　：（目線Ｘ・Ｙ・Ｚ軸）×（平均・分散・最低・最高）
  - 頭角度　　：（ピッチ・ロール・ヨー）×（平均・分散・最低・最高）
  - 表情検出座標　：（横幅・高さ）×（平均・分散・最低・最高）
    - このデータは間接的にカメラに対する顔の近さを判定するデータとして使う。
  - 口状態　　：（閉口・開口・他）×（平均・分散・最低・最高）
  - 静的データ　　：（推定年齢×最頻値）＋性別＋メガネ

続いて、無効なデータの排除を行った。具体的には動画フレームに顔が写っていない等の理由で Face++による分析が行えなかった表情データや、アンケートが未回答のデータ等である。これらを排除した結果表情データとアンケートデータの組を「難しさ：453 個」「面白さ：453 個」「疲労：452 個」「集中度：285 個」の有効なデータを得た。

次に、圧縮後のデータをそのままニューラルネットに用いた場合、ニューラルネットで使用する活性化関数の関係上、極端な値の振れを起こしてニューラルネット内の各ノードの振る舞いが単純パーセプトロンのような動きになり学習を阻害する恐れがある。そのため、圧縮後のデータにおける各系列に対してその値の範囲での正規化を行った。この例を以下の図に示す。

図 5-1 データ正規化の例

図 5-1 の例では表情成分：悲の平均 453 個の系列に対して、系列内の最高値と最低値を求め、これを基準に正規化している。

第5.3節 使用したライブラリに関して

本章においてニューラルネットワークによる学習を行うライブラリとして TensorFlow [16]と Keras [17]を使用した。

TensorFlow は元々ベクトル計算を専門としたライブラリであるが、ニューラルネットを扱うことを意識した関数が多く、ニューラルネットワークの設計において高い自由度を持つことが特徴である。選定した理由は本論文執筆時点で機械学習分野における利用事例が多い事や資料の存在、フォーラムによる意見交換が活発であることが挙げられる。また、後述の Keras がサポートしているライブラリである点も大きい。

Keras はそれ単体が機械学習を行うライブラリではなく、TensorFlow や Theano といった別のライブラリ上で動作するもので、下位のライブラリに渡すコードが関数としてまとめてあるので、ニューラルネットワークを構成する上で簡便な記述で済むことや、自然言語処理や画像分析といった特定分野を意識した関数を多数備えている事から汎用性が高いという理由で採用した。

第5.4節 予備実験

本節では第 4 章にて得られたデータに対しニューラルネットワークを利用して解析・推定するにあたり、いかなるネットワーク構造が解析や予想に適しているかを確認するために行った実験について記述する。

ここでは第 3 章で得られたデータの内、表情のデータと難易度に関するデータの組み合わせのみを用いた。この組み合わせ 453 個を無作為に選んだ 30 個のテスト用データと残り 423 個の訓練用データに分け、様々なニューラルネットワークに対して、訓練用データを用いてバッチ学習を行わせ、学習のエポックごとにテストデータを入出力として予測し、その正解率を記録した。

この作業において抜き出すデータを 50 回変更し、各エポックの正解率の平均値を指標として比較・検討を行った。

第5.4.1項　隠れ層数別の比較

本項では以下のような条件でニューラルネットワークの学習を行った。

- 入力層のノードは 83 個（入力データの次元数と同じ）
- 隠れ層はシグモイド関数を用いノードは 83 個（入力データの次元数と同じ）
- 出力層はソフトマックス関数を用い、ノードは 5 個
- 学習率は 0.2
- 隠れ層の層数を 2 ～ 9 層に変更し、比較する。

以上の条件で学習を行った。エポックごとの正解率の比較と推移を以下のグラフに示す。（縦軸が正解率、横軸がエポック数である）

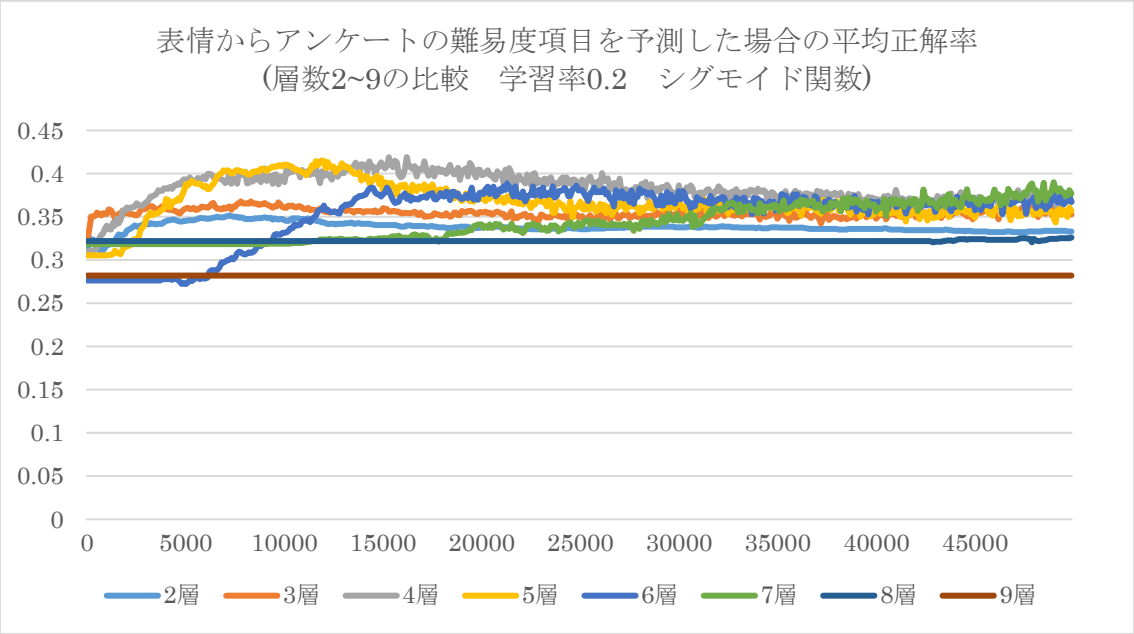

図 5-2 学習率 0.2、シグモイド関数使用時の層数別正解率比較

層数 2 ～ 9 の全てで平均正解率が半分を切っており、平均正解率の最高は層数 5 層でエポック 12200 回目の正解率 0.41 であった。

グラフを見るあたり、2~5 層の平均正解率の推移は学習開始から徐々に上がり、エポック 15000~20000 回をピークにだんだん減少し、平均正解率 0.35 付近で収まっている。これはエポック数が多すぎると過学習（訓練用データに過剰に適応し汎用性を失う現象） [13]によってテストデータに適応できなくなっているものと思われる。

しかし 7 層ではピークが遅く、 8 ～ 9 層に至ってはこの実験で行ったエポック 50000 回の間は全く正解率が変化していない。これは層数をこれ以上増やしても正解率が増減しない学習の限界なのか、単にエポック数が不足していたものなのかを確認したほうが良いと考えられるため、次項の実験を行った。

第5.4.2項 高学習率における隠れ層数別の比較

本項では以下のような条件でニューラルネットワークの学習を行った。
第 5.4.1 項との相違点を太字で示す。

- 入力層のノードは 83 個（入力データの次元数と同じ）
- 隠れ層はシグモイド関数を用いノードは 83 個（入力データの次元数と同じ）

- 出力層はソフトマックス関数を用い、ノードは 5 個
- 学習率は 5.0 および 10.0
- 隠れ層の層数を 6～１０層に変更し、比較する。

以上の条件で学習を行った。エポックごとの正解率の比較と推移を以下のグラフに示す。(縦軸が正解率、横軸がエポック数である)

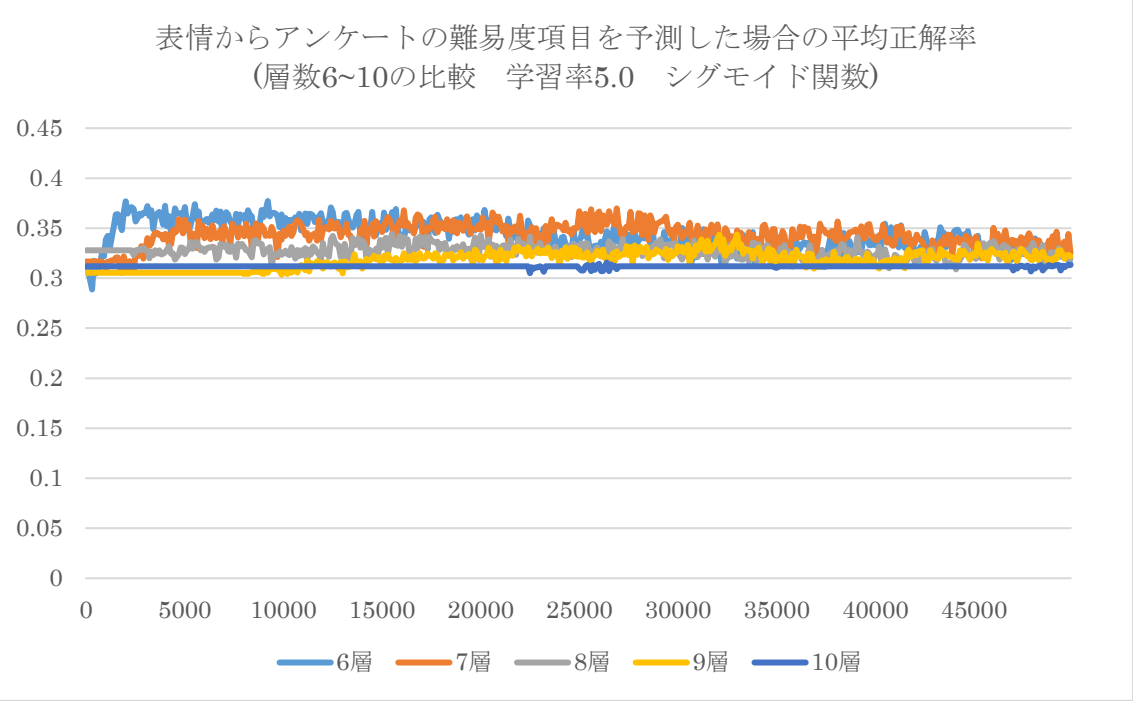

図 5-3 学習率 5.0、シグモイド関数使用時の層数別正解率比較

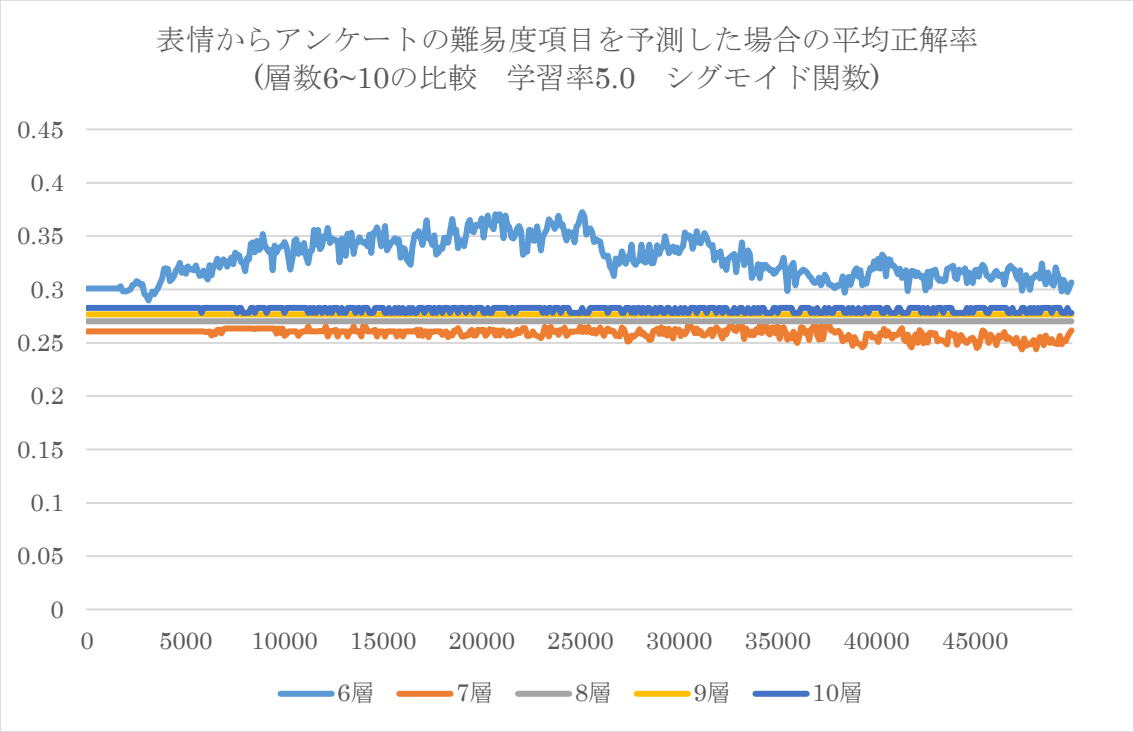

図 5-4 学習率 10.0、シグモイド関数使用時の層数別正解率比較

結果、いずれの学習率、いずれの層数においても第 5.4.1 項で得られた平均正解率を下回ることから、層数を増やしても平均正解率は上昇しないものと結論した。

第5.4.3項 活性化関数別の比較

第 5.4.1 項で得られた平均正解率でも低いため、他に平均正解率を上げる手段として、ニューラルネットワークの活性化関数を変更して学習を行い、その平均正解率と第 5.4.1 項の平均正解率と比較することとした。

ここで用いる活性化関数として双曲線正接関数(tanh 関数)とランプ関数(ReLU 関数)を挙げる。tanh 関数は導関数がシグモイド関数の導関数と比べて 0 近傍の値が大きいため、勾配消失問題（隠れ層の層数もしくはノードを増やすと不正解時に行われる重みの勾配（修正値）が微小になってしまう現象）を起こしにくいとされる。また、ReLU 関数は 0 以上の変数に対して単調増加であり、導関数は 0 以上の変数に対して常に 1（導関数がステップ関数と同一）であるため、高次元・重層においても勾配消失を起こさない、しかし 0 未満に対しては導

関数 0 なので学習中に不活性になったノードは再活性化しないという問題もあるとされる [13]。

本項では以下のような条件でニューラルネットワークの学習を行った。

- 入力層のノードは 83 個（入力データの次元数と同じ）
- 隠れ層は**双曲線正接関数(tanh 関数)**及び**ランプ関数(ReLU 関数)**を用いノードは 83 個（入力データの次元数と同じ）
- 出力層はソフトマックス関数を用い、ノードは 5 個
- 学習率は 0.2
- 隠れ層の層数を 2～6 層に変更し、比較する。

以上の条件で学習を行った。エポックごとの正解率の比較と推移を以下のグラフに示す。（縦軸が正解率、横軸がエポック数である）

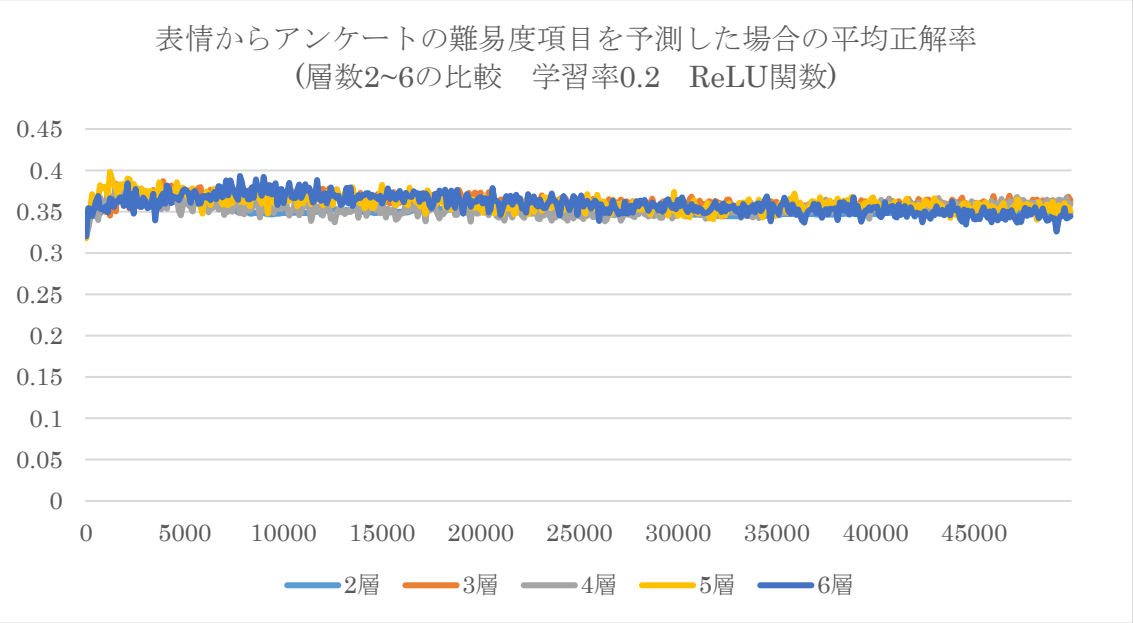

図 5-5 学習率 0.2、tanh 関数使用時の層数別正解率比較

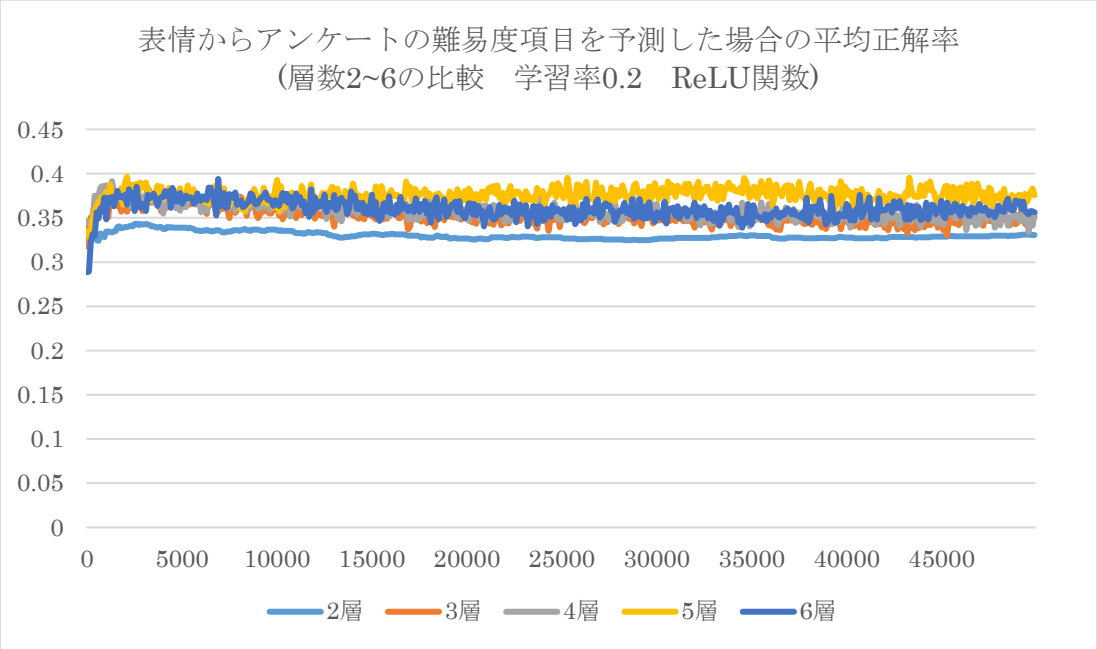

図 5-6 学習率 0.2、ReLU 関数使用時の層数別正解率比較

シグモイド関数以外の平均正解率の推移を見ると、tanh 関数・ReLU 関数ともに正解率の向上が早く、過学習による正解率低下を起こしにくい事が確認できたが、いずれにしても平均正解率の最高値はシグモイド関数を使用した場合と同程度かそれ以下に落ち着くものと結論した。

第5.4.4項 質の良いデータの選抜した場合の比較

第 5.4.1 項から第 5.4.3 項までの予備実験ではいずれも第 4 章で得られた難易度のデータの全て使用してきた。しかし、このデータには一部に質の悪いデータがある可能性がある。

具体的にはある調査協力者は理数系が苦手だという理由で難易度の項目全てを 5（難しい）の判定をつけたり、実験前にアルバイトをしていたため実験開始時点で疲労していたから疲労の項目を全部 5（疲れている）の判定をつけたりするなどのケースが見られた。それ自体は協力者の主観として正しいのだが、ニューラルネットワークに学習させるにあたって「表情が変化しているのに正解が変わらないデータが、不正解時にネットワーク全体の勾配（修正値）を下げている」可能性がある。

そのため、全てのアンケート結果のうち、各協力者の回答のばらつき具合で質を判定し、良質なデータのみ用いた場合、推定結果の平均正解率がどの様になるかを分析した。

この時の質の判定を以下の表に示す。

表 5-1 アンケート結果の質の判定

協力者	回答数	難易度 d	面白さ i	疲労 t	集中度 c	備考
1	30	B	C	B	未収集	
2	30	A	A	B	未収集	
3	30	A	A	A	未収集	
4	30	C	C	A	未収集	
5	30	B	B	B	未収集	
6	24	A	A	A	未収集	
7	21	A	A	A	未収集	
8	29	A	A	A	A	
9	21	A	A	A	A	
10	19	A	A	A	A	
11	30	A	B	C	C	回答ミス 1 問あり
12	28	A	A	A	B	未回答 1 問あり
13	19	A	A	A	A	
14	23	A	A	B	B	
15	30	A	C	C	C	
16	30	B	無回答	C	C	
17	19	B	A	B	A	
18	27	B	C	C	C	
19	18	A	A	A	B	
A＝回答が 4 ～ 5 種類、B＝回答が 3 種類、C＝回答が 1 ～ 2 種類						

この表の判定結果を元に、質の良いデータのみで学習した場合どうなるかを確認しようと考えた。以上の理由で本項では以下のような条件でニューラルネットワークの学習を行った。

- 入力層のノードは 83 個（入力データの次元数と同じ）
- 隠れ層はシグモイド関数を用いノードは 83 個（入力データの次元数と同じ）
- 出力層はソフトマックス関数を用い、ノードは 5 個
- 学習率は 0.2
- 隠れ層の層数を 2 ～ 6 層に変更し、比較する。

- **使用するデータは表 5-1 の内、AB 判定のみを用いる場合と、A 判定のみ用いる場合に分ける。**

以上の条件で学習を行った。エポックごとの正解率の比較と推移を以下のグラフに示す。(縦軸が正解率、横軸がエポック数である)

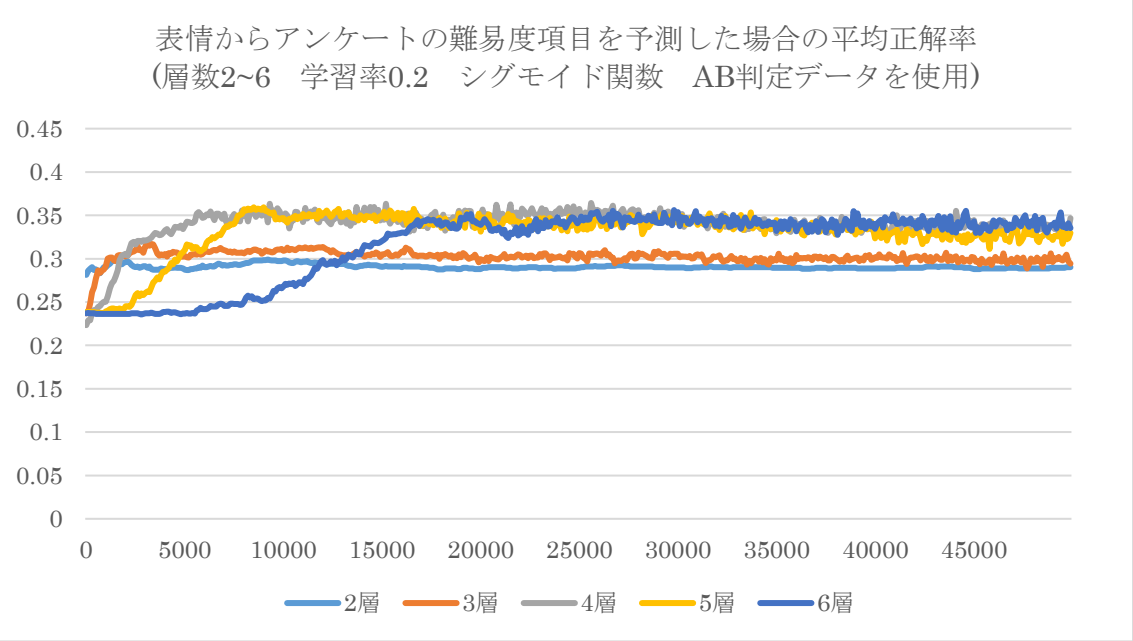

図 5-7 AB 判定データ使用時の層数別正解率比較

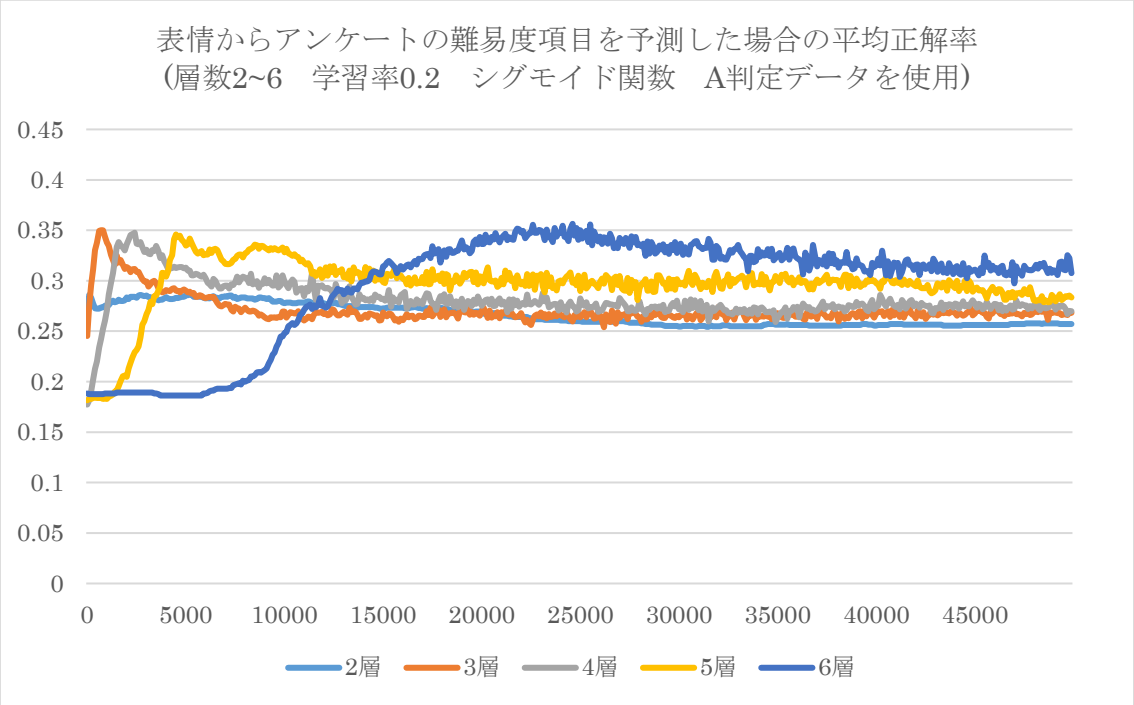

図 5-8 A 判定データ使用時の層数別正解率比較

AB 判定のみを使用した場合、２～３層において平均正解率の伸びが悪く、４層以降は層数が増えるにつれて平均正解率のピークが早くなっているが、最終的に第 5.4.1 項の正解率を超えることはなかった。

また、A 判定のみを使用した場合、6 層以外は平均正解率のピーク後、過学習により急激な平均正解率の低下が見られること（特に 3 層が顕著である）、また層数が増えるにつれピーク後に平均正解率の停滞した時の値が高くなっている点が興味深いが、最終的に第 4.4.1 項の正解率を超えることはなかった。

考察としては、もともとのデータ数が少ない（難易度データは ABC 全部で４５３個、AB 判定で４２３個、A 判定のみで２９７個）事や、A・AB いずれのケースもテスト用データの数が３０個で固定であったため、データ全体に占める訓練用データの割合が少なくなったことも原因の可能性がある。

第5.4.5項 正規化方法を変更した場合の比較

第 4.4.1 項から第 5.4.4 項までの予備実験ではいずれも第 5.2 節で述べた方法で正規化されたデータを用い学習を行ったが、本来 Face++から渡される数値はある程度の正規化がなされている。例えば表情成分や目線の分析値は全て 0 ～

１００の範囲に収まるようになっており、頭部角度の値は全て－１８０～１８０の範囲に収まるようになっている。この Face＋＋の値範囲の仕様を用いて正規化を行ったデータで学習すると平均正解率がどのようになるかを観察する。

この正規化方法の差を以下に示す。

- 第 5.2 節に記した正規化
  ある系列における値が「0.0，5.0，10.0，15.0，20.0」なら系列内の最低値が 0、最高値が 20 なので、正規化後の系列は「0.0，0.25，0.5，0.75，1.0」となる。
- 本項で述べる正規化
  ある系列における値が「0.0，5.0，10.0，15.0，20.0」で、Face＋＋から渡される値の範囲が 0～100 の場合、
  正規化後の系列は「0.0，0.05，0.1，0.15，0.2」となる。

上記の２つの方法のうち、後者を用いて再度データの正規化を行い、これを元に学習を行う。よって、本項では以下のような条件でニューラルネットワークの学習を行った。

- 入力層のノードは 83 個（入力データの次元数と同じ）
- 隠れ層はシグモイド関数を用いノードは 83 個（入力データの次元数と同じ）
- 出力層はソフトマックス関数を用い、ノードは 5 個
- 学習率は 0.2
- 隠れ層の層数を２～６層に変更し、比較する。
- **正規化の基準を Face++の解析値の仕様範囲に合わせたデータで学習する。**

以上の条件で学習を行った。エポックごとの正解率の比較と推移を以下のグラフに示す。（縦軸が正解率、横軸がエポック数である）

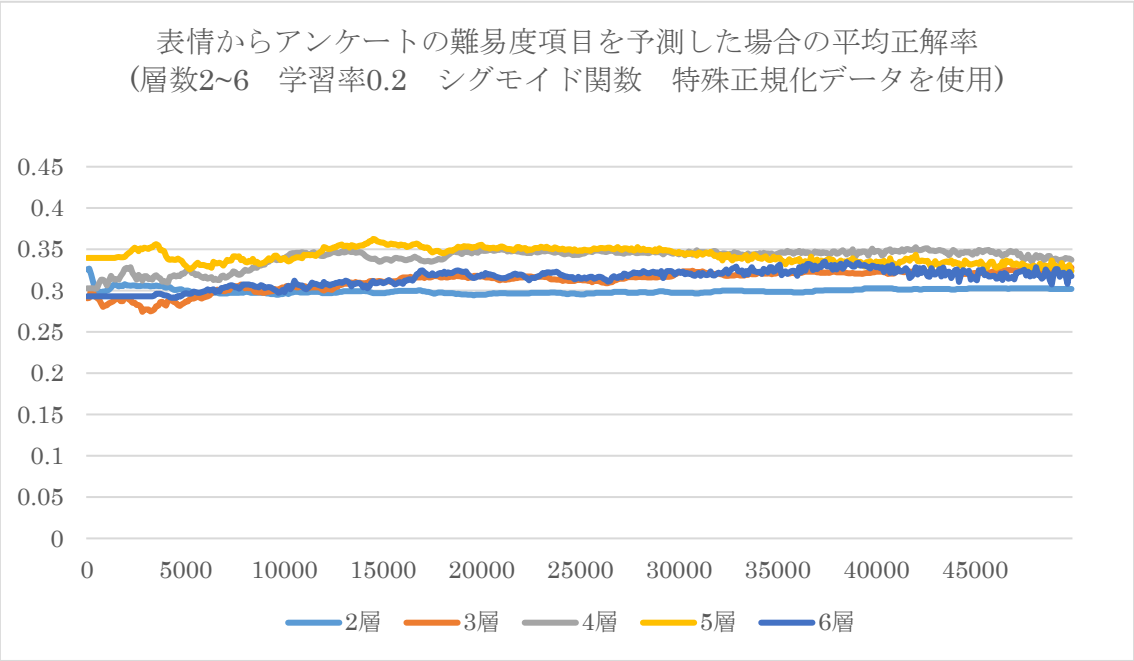

図 5-9 正規化方法を変更したデータ使用時の層数別正解率比較

結果、すべての層数において第 5.4.1 項の平均正解率を下回る結果となった。また、すべての層数においてピーク時とそうでない時の区別がつきにくいことや、過学習による平均正解率の低下があまりないといった特徴がみられる。

考察として、正規化方法の変更により第 5.2 節で述べた正規化方法と比べて全ての表情データの値が下がったため、不正解時の勾配（修正値）が低くなった結果、モデルが学習しにくくなり、アンダーフッティング [18]（粗学習、ニューラルネットワークが訓練データを近似できていない。過学習やオーバーフッティングの逆）を起こしやすくなったことが、学習の前後においても平均正解率があまり変わらない原因ではないかと考えられる。

第5.4.6項 予備実験の結論

第 5.4.1 項における平均正解率が当初低いと考えられ、これを改善するために様々な方法を試したものの、平均正解率を改善することはできなかった。よって、第 5.4.1 項のネットワーク構造を他のデータを解析するのに適当と結論し、次章において他のデータの解析に使用する。

第5.5節 本実験

本節では第 4 章において入手したデータを第 5.4.6 項にて結論したニューラルネットワークを用いて学習を行い、その結果の概観と考察を行う。

本節に記載される学習結果は全て以下の条件によって行われた。

- 入力層のノードは 83 個（入力データの次元数と同じ）
- 隠れ層はシグモイド関数を用いノードは 83 個（入力データの次元数と同じ）
- 出力層はソフトマックス関数を用い、ノードは 5 個
- 学習率は 0.2
- 隠れ層の層数を 2 ～ 7 層に変更し、比較する。

第5.5.1項 「難しさ」の項目の学習結果

図 5-10 は第 5.4.1 項と数値は同じで変形しただけである。

基本的な概要は第 5.4.1 項で述べたとおりで、全体的な正解率が低い。また、0～5 層のいずれについてもエポック数が進むに連れて 0.35 付近で平均正解率が収束していること、0 層と 5 層を除いて過学習の影響による平均正解率の低下が見られること（ただし 5 層は過学習に至る学習回数まで足りないだけである可能性がある）等が見られるが、いずれにしても正解率は低い。

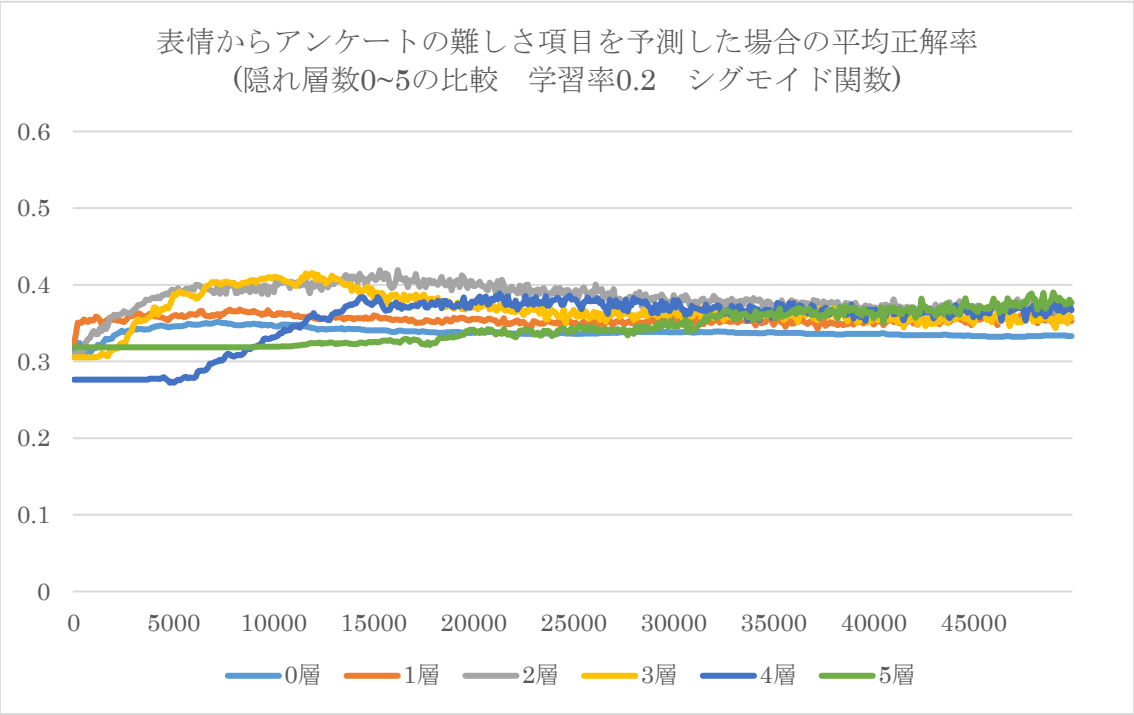

図 5-10 難しさの項目の層別平均正解率

第5.5.2項 「面白さ」の項目の学習結果

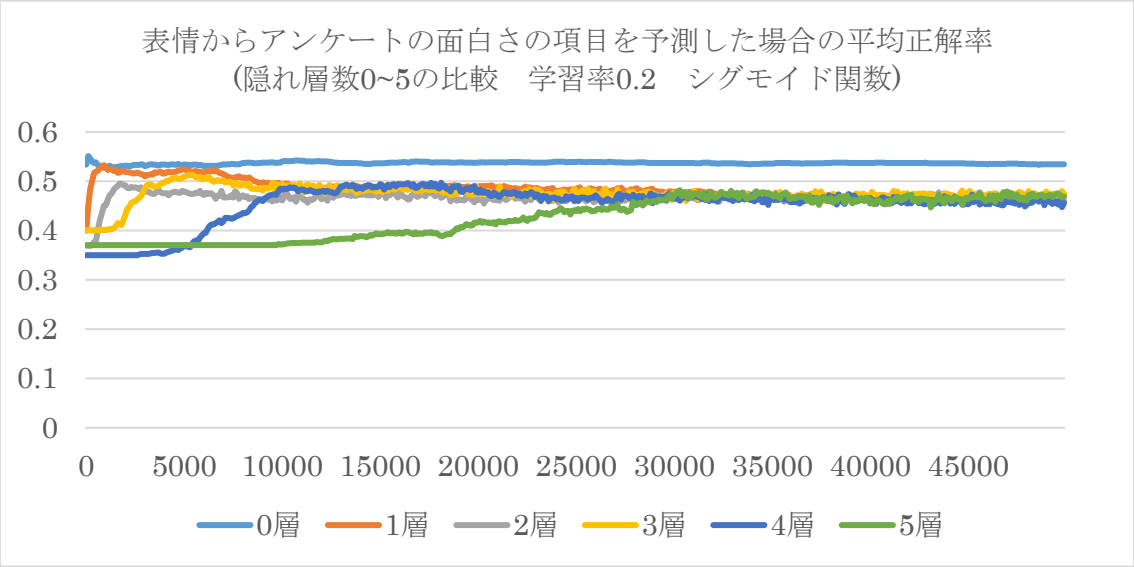

図 5-11 面白さの項目の層別平均正解率

前項において難しさの項目の平均正解率が低かったことから他の項目についても低いと考えられていたが、この面白さの項目では 0~1 層において最高値 0.55 と高い平均正解率を記録している。特に 0 層という実質的にソフトマックス関数の層しか存在しない状態が最も平均正解率が高く、しかもエポック数を重ねても正解率が低下しないことを考えると、面白さと表情のデータの関連性は単純なものである可能性が考えられる。

第5.5.3項 「疲労」の項目の学習結果

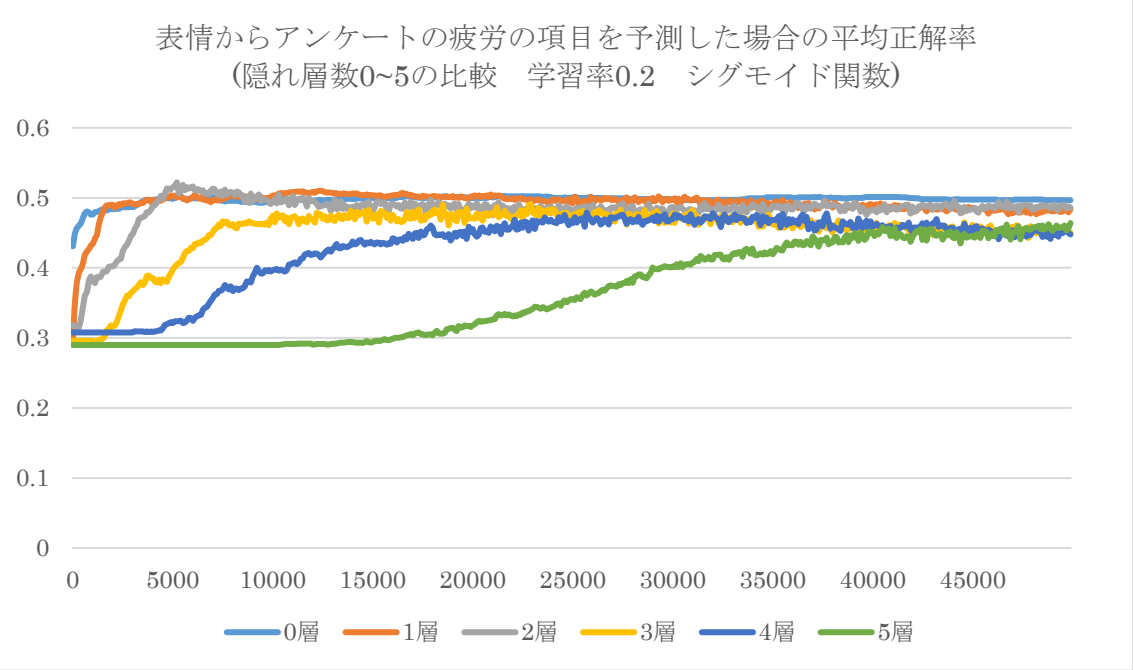

図 5-12 疲労の項目の層別平均正解率

疲労の項目の平均正解率において最高値は 2 層の 0.522 であった。最高は 2 層であるが、グラフの推移を見るとエポックを重ねても最も正解率が低下しないのは 0 層である。面白さの項目においても 0 層目がエポック数を重ねても平均正解率が低下しなかったことから、何らかの関係性があると考えられる。

第5.5.4項 「集中度」の項目の学習結果

集中度の項目における平均正解率の最高値は 0 層の 0.505 であった。0 層がエポックを重ねても平均正解率が下がらない現象がまた発生している。 0 ～ 2 層までは平均正解率のピーク後にやや下がる傾向が見えるが 3 ～ 5 層は比例ではな

いが単調な増加に見える。また、エポックを重ねた際の平均正解率（45000～50000 回あたり）について、0 層以外ある程度収束する傾向が見られる。

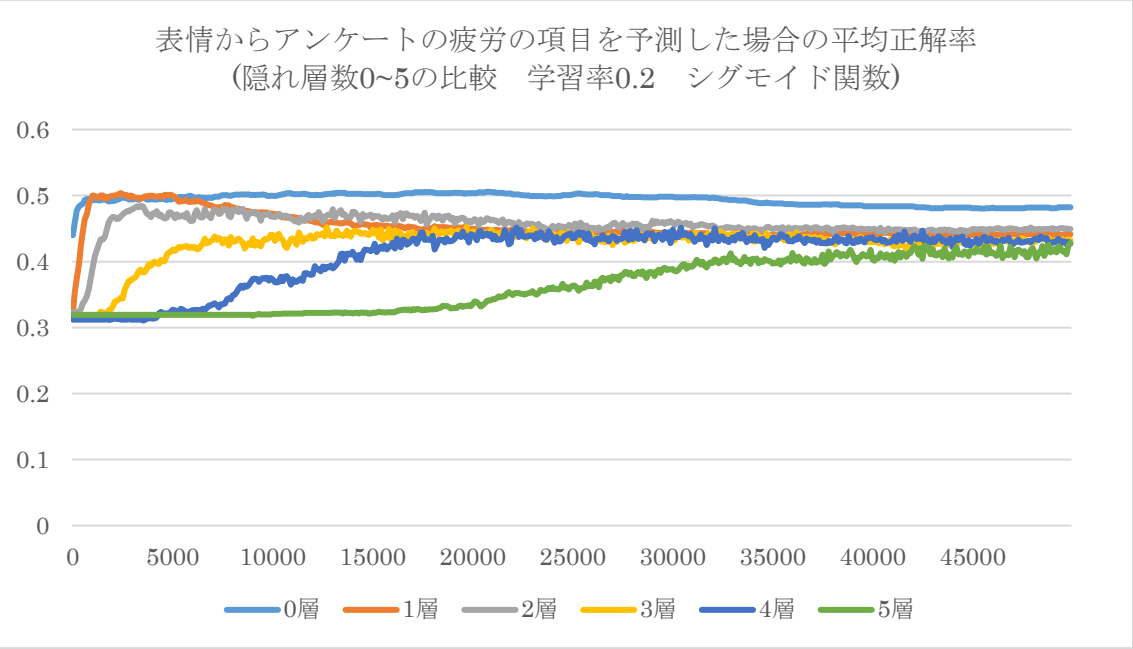

図 5-13 集中度の項目の層別平均正解率

第5.5.5項 総括

アンケートの全ての項目に対する平均正解率を見ると、その推移のパターンを「難しさの項目」と「それ以外の 3 項目」に分けることができる。難しさの項目では全体的な平均正解率 0.3～0.4 の範囲に収まり、0 層の正解率もさほど高くない。

それ以外の 3 項目は全体的な平均正解率が 0.4~0.5 の範囲に収まり、共通して 0 層の正解率があまり下がらないという特徴が見える。ただし 3 項目の中でエポック 45000~50000 の範囲において各層数ごとの平均正解率が 0 層以外収束する様子が見えるが、この収束のばらつき具合が「疲労＞集中度＞面白さ」の順になっている。また、集中度の項目のみ 30000 エポック付近で平均正解率がやや下がっているが、今までの 0 層の平均正解率の特異性から考えるとこれが過学習によるものなのか断定できない。

このようなグラフになった原因について推測としては

- 難しさ以外の 3 項目は表情との関連が単純である可能性。

- この場合、難しさの項目だけ他の比べ関連が複雑と思われる。
- もしくは難しさの項目だけ表情との関連性がなく、十分な推定が行えていなかった可能性がある。

- 難しさの項目だけ抜き出したテストデータが偏っていた可能性。
  - そういった偏りを防ぐために各層で 50 回テストデータを抜き変えているがそれでも不十分なのか。
  - もしくはアンケートの不備により調査協力者ごとの主観の差によって違う関連性が混在したか。
- データ量が少なすぎるが故に発生した誤差の大きさの問題であって、各項目の関連性の問題ではない可能性
- テストデータと訓練データの比率が悪い可能性。

等が考えられるが、現時点では考察の材料が少なく結論できない。

第6章 他の分析手法の検討

本章では第 4.1.1 項で述べたニューラルネットワーク以外にもアンケート結果の予測に使用できる手法やその是非について論じる。

第6.1節 統計的手法による分析

本節では第 5.2 節にて述べられた加工・正規化された表情データとアンケートの結果に関連性があるかどうかを統計的手法によって分析した結果を述べる。

第6.1.1項　相関関係

本項では第 5.2 節にて述べられた加工・正規化された表情データの各系列に対して、excel を用いてアンケート結果の数値との相関係数を計算した結果、およびそれに対する考察を述べる。

以下に、計算した相関係数の表を記載する。

表 6-1 表情データとアンケートデータの相関係数表

			相関	相関	相関	相関
			d_難しさ	i_面白さ	t_疲労	c_集中度
emothion	sadness	avg	-0.06514	-0.00479	-0.01269	0.019453
		var	-0.05734	0.034405	0.002828	-0.04981
		max	-0.03261	0.090218	-0.02089	-0.08388
		min	0.012445	-0.03	-0.00098	0.172746
	netural	avg	0.132968	-0.23822	0.157255	0.092153
		var	-0.01383	-0.09453	-0.00748	0.059596
		max	0.069145	-0.06545	0.129141	-0.0758
		min	0.130003	-0.16197	0.13759	0.11816
	disgust	avg	0.083064	0.088166	0.131878	-0.25376
		var	-0.01678	-0.04329	-0.03613	-0.01784
		max	-0.00331	0.105741	-0.01381	-0.12208
		min	0.06833	0.148542	0.106954	-0.06925
	anger	avg	-0.15263	0.014282	-0.27994	0.21528
		var	-0.07877	0.068749	-0.09705	0.068244
		max	-0.16791	0.003667	-0.21333	-0.00194
		min	-0.04206	-0.0575	-0.04957	0.112693
	surprise	avg	0.011435	-0.04928	0.048709	-0.00517

		var	0.005294	-0.05091	0.017437	-0.02671
		max	0.075893	0.099892	0.117262	-0.06024
		min	0.051504	-0.01906	0.038318	0.112898
	fear	avg	0.046488	-0.01264	0.125402	0.127989
		var	0.033405	0.013849	0.112223	0.045843
		max	0.032616	0.035639	0.136838	0.003147
		min	0.048067	-0.02636	0.05639	0.115366
	happiness	avg	-0.01259	0.06175	0.115614	0.056767
		var	0.011397	0.085827	0.134561	0.02534
		max	0.00067	0.010779	0.173053	-0.06055
		min	0.052278	-0.02187	0.047431	0.176308
eyegaze	right_eye_x	avg	0.131632	-0.20163	0.042429	-0.24433
		var	0.290922	-0.36593	0.095488	0.08488
		max	0.198162	-0.15658	0.143013	-0.33885
		min	-0.10961	0.126854	0.011472	0.125081
	right_eye_y	avg	-0.01668	-0.16169	-0.10694	0.262094
		var	0.212445	-0.14235	0.156764	0.029334
		max	0.01006	-0.12208	-0.15985	0.173532
		min	-0.14613	-0.03387	-0.06012	0.309898
	right_eye_z	avg	-0.18676	0.308803	-0.09285	-0.41779
		var	0.261675	-0.34426	0.108852	0.190465
		max	0.024297	0.100963	-0.06866	-0.40789
		min	-0.06161	0.179438	0.025575	-0.02074
	left_eye_x	avg	0.242078	-0.35978	0.166227	0.241398
		var	0.142933	-0.17451	0.105682	0.093637
		max	0.272765	-0.1824	0.226687	0.111658
		min	0.053196	-0.01099	0.083977	0.197625
	left_eye_y	avg	0.027887	-0.13849	-0.00187	0.013774
		var	0.074315	0.056174	0.000373	-0.02181
		max	-0.07644	-0.05797	-0.16645	0.027236
		min	-0.03477	-0.17835	0.063358	0.206008
	left_eye_z	avg	-0.26314	0.388069	-0.17145	-0.20909
		var	0.239236	-0.32143	0.148155	0.239081
		max	-0.11687	0.266138	-0.12311	-0.23938
		min	-0.1734	0.121151	-0.02724	-0.02703

retangle	width	avg	-0.02984	-0.20472	0.034351	0.055557
		var	0.014773	0.043576	0.047143	0.087792
		max	-0.03124	-0.14406	0.028951	0.061671
		min	-0.01687	-0.2536	0.078898	0.065989
	height	avg	-0.02984	-0.20472	0.034351	0.055557
		var	0.014773	0.043576	0.047143	0.087792
		max	-0.03124	-0.14406	0.028951	0.061671
		min	-0.01687	-0.2536	0.078898	0.065989
headpose	pitch	avg	-0.1464	-0.10725	-0.05231	0.267344
		var	0.042389	-0.04144	0.11046	0.043567
		max	-0.10267	0.044343	-0.11381	0.192248
		min	-0.15394	-0.08995	-0.14922	0.192806
	roll	avg	-0.07652	0.086814	-0.07191	0.027159
		var	0.079402	-0.09926	0.137015	0.11337
		max	0.036436	0.065246	0.064782	0.045703
		min	-0.11285	0.058303	-0.08192	0.016164
	yaw	avg	0.134242	-0.12849	-0.00503	-0.06556
		var	0.048917	-0.05741	0.045913	0.018568
		max	0.035771	0.056128	-0.22156	-0.06631
		min	0.080155	-0.12583	-0.01811	0.008476
mouth	close	avg	0.194976	0.00755	0.092389	-0.03394
		var	-0.07423	0.025532	0.013108	-0.06694
		max	0.089503	0.030784	0.03973	0.058397
		min	0.103031	-0.05646	0.08306	-0.03211
	oppen	avg	-0.18187	-0.05087	-0.12475	0.198518
		var	-0.12023	0.021539	-0.02523	0.105071
		max	-0.11264	0.060442	-0.11529	0.01038
		min	-0.0327	-0.07054	-0.00824	0.05492
	other	avg	-0.08721	0.044898	-0.01407	-0.11751
		var	-0.07029	0.037521	-0.02247	-0.18547
		max	-0.06535	0.131263	-0.10671	0.020978
		min	-0.08215	-0.01805	-0.03507	-0.06634
age	age	age	0.07853	-0.05043	0.09246	-0.27634
gender	gender	gender	0.022109	0.251817	0.07239	-0.21362
glass	glass	glass	0.331648	0.129514	0.349083	-0.05271

なお、表 6-1 には相関係数が 0.2 以上のセルには桃色、-0.2 以下のセルには青色で着色している。

また、この表において、相関係数の絶対値の高いセルの個数と、表内の相関係数の最低値・最高値に関してまとめた表を以下に示す。

表 6-2 表 6-1 における相関係数の集計

	相関絶対値の範囲の個数				相関絶対値の積算個数(上側)			
	難しさ d	面白さ i	疲労 t	集中度 c	難しさ d	面白さ i	疲労 t	集中度 c
0.2 ≦|r|< 0.21	0	3	0	2	8	14	5	16
0.21 ≦|r|< 0.22	1	0	1	2	8	11	5	14
0.22 ≦|r|< 0.23	0	0	2	0	7	11	4	12
0.23 ≦|r|< 0.24	1	1	0	2	7	11	2	12
0.24 ≦|r|< 0.25	1	0	0	2	6	10	2	10
0.25 ≦|r|< 0.26	0	3	0	1	5	10	2	8
0.26 ≦|r|< 0.27	2	1	0	2	5	7	2	7
0.27 ≦|r|< 0.28	1	0	1	1	3	6	2	5
0.28 ≦|r|< 0.29	0	0	0	0	2	6	1	4
0.29 ≦|r|< 0.3	1	0	0	0	2	6	1	4
0.3 ≦|r|< 0.31	0	1	0	1	1	6	1	4
0.31 ≦|r|< 0.32	0	0	0	0	1	5	1	3
0.32 ≦|r|< 0.33	0	1	0	0	1	5	1	3
0.33 ≦|r|< 0.34	1	0	0	1	1	4	1	3
0.34 ≦|r|< 0.35	0	1	1	0	0	4	1	2
0.35 ≦|r|< 0.36	0	1	0	0	0	3	0	2
0.36 ≦|r|< 0.37	0	1	0	0	0	2	0	2
0.37 ≦|r|< 0.38	0	0	0	0	0	1	0	2
0.38 ≦|r|< 0.39	0	1	0	0	0	1	0	2
0.39 ≦|r|< 0.4	0	0	0	0	0	0	0	2
0.4 ≦|r|< 0.41	0	0	0	1	0	0	0	2
0.41 ≦|r|< 0.42	0	0	0	1	0	0	0	1
0.42 ≦|r|< 0.43	0	0	0	0	0	0	0	0

表 6-2 の表を見ると、難しさと疲労の項目で相関のある項目が少ない事から、表情データとアンケート項目の相関性が少ないことが予想されるが、もしそうなら第 5.5.1 項と第 5.5.3 項でニューラルネットワークによるアンケート結果の推定で正解率に差がついたのかを説明できない。しかも相関係数が高いセルの数では疲労の項目の方が少なく、相関係数の最低・最高値を比較しても値の差は大きくない。

仮説としては、疲労の項目の方に線形な相関以外の相関（2 次相関や円相関など）が存在し、それをニューラルネットがうまく近似した結果、難しさの項目と平均正解率に差が出た可能性があり、さらなる検証が必要と結論する。

第6.2節 RNN による分析の可能性

RNN(Recurrent Neural Network・再帰ニューラルネットワーク)とは隠れ層において、過去に入力されたデータを記憶し、不正解時の勾配（修正量）の計算を遡って行うことで時系列的なデータの予測・分類が可能なニューラルネットワークの一種である [13]。

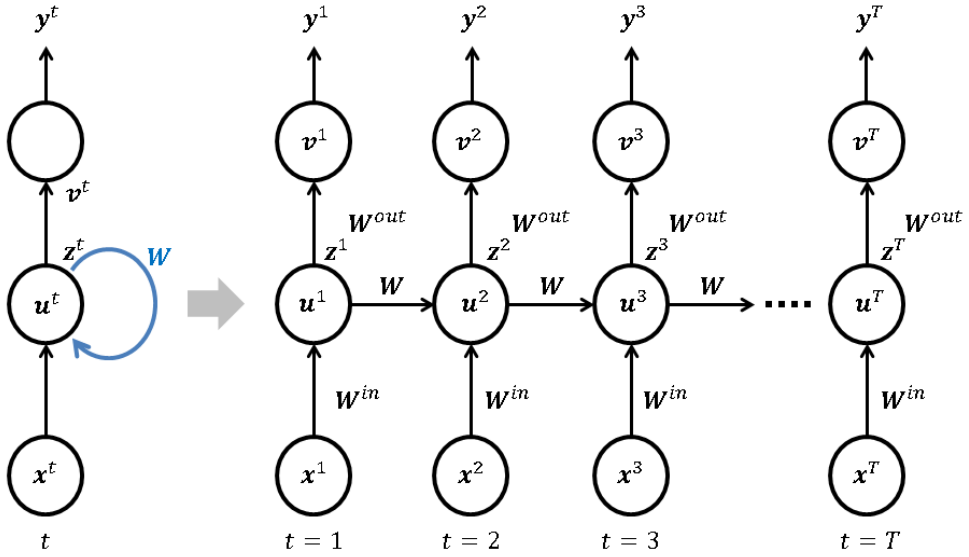

図 6-1 RNN の概念図 [14]

RNN の特徴として「時系列データの分類ができる」事と「入出力の数が一致していなくても良い」という特徴がある。この特徴は自然言語処理において文章を入力して分類や続く文書の生成をさせたり、音声解析において各周波帯

の音の大きさを入力として音声認識や続く音声の予想をさせたりするなどの形で利用されている。

この特徴から、RNN に対して第 5.2 節で加工される前のデータ（第 4.3 節で例示したデータ）が動画のフレームの連なりという時系列なデータであるため、これをそのまま RNN に入力することでアンケート結果を予想することができると思われるため、これの利用を今後の課題としたい。

第6.3節 CNN による分析の可能性

CNN(Convolutional Neural Network・畳み込みニューラルネットワーク)とはニューラルネットワークにおいて、入力の次元数を一定の範囲で区切り（プーリング・刈り込み)、その範囲内でニューラルネットをさらに学習・出力させることで扱うデータの次元数を削減しながら計算してくニューラルネットワークの一種である。

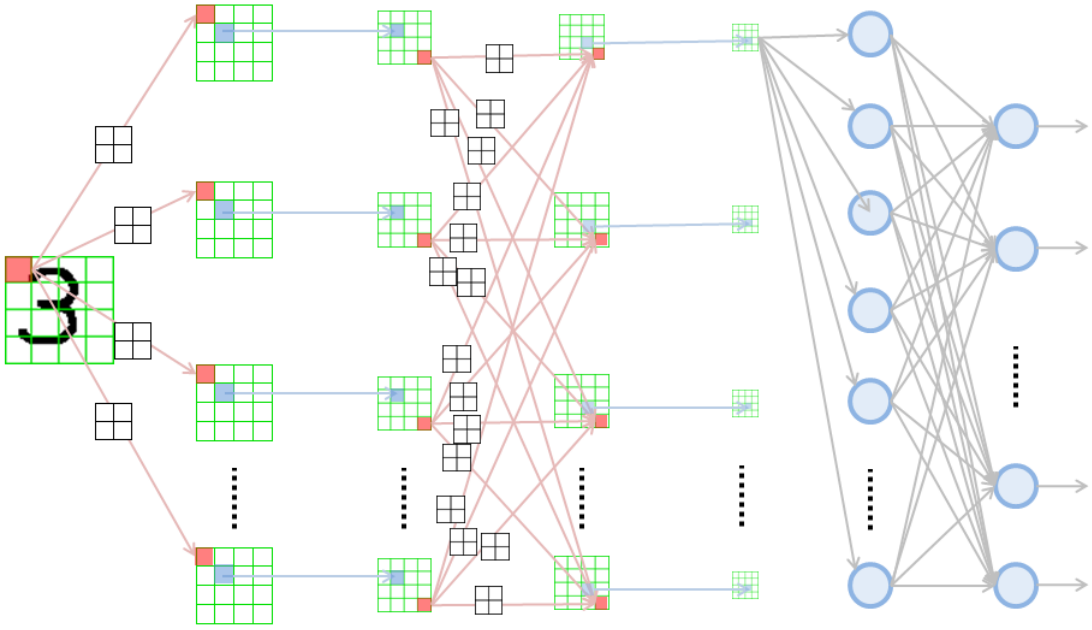

図 6-2 CNN の概念図 [14]

CNN は google のチームが事前にニューラルネットに判断基準となるデータを与えずに猫とそれ以外の画像識別に成功 [15]して以降、分類の基準となる特徴をモデル自身が探すことができるという利点から利用されている。例えば大

学教育において論文が文の体をなしているかどうかを判定するソフトウェアが利用されていたりする。

本来 CNN は画像処理を目的としているが、入力データの次元ごとの順番に意味のあるデータ（ある値の系列を入れ替えると意味のなくなるデータ、上記の図 6-2 は画像のピクセルを入力として、1 段目と 2 段目のデータを入れ替えると別の画像になってしまう）であれば何にでも利用ができるので、第 5.2 節で加工する前のデータであれば利用できるのではないかと考えられるため、今後の課題としたい。

第7章 おわりに

第7.1節 結論

本研究では第 1.1.3 項において述べた「学習中の生徒の表情のデータから学習に関する心理的な情報、所謂モチベーションのパラメータを推定することが出来るかどうかの検証を行うこと」という目的を達成するため、学習中のモチベーションと表情には何らかの関係性があるという仮定の元、第 3 章の予備調査において課題を洗い出しながら、第 4 章において調査協力者 19 名に対し、CAB 模擬問題を PC で解いてもらい、その作業中の表情と PC の画面を動画で撮影した。CAB 模擬問題の回答終了後に各問題に対しての「難しさ」「面白さ」「疲労」「集中度」といったモチベーションに関係・影響すると思われる 4 つの項目についてアンケートを取り回答を得た。

このようにして得た表情のデータとアンケートのデータに対し、ニューラルネットワークを用いて表情のデータからアンケートのデータを推定することができるかを実験するため、データを加工した上で実験を行った。

結果、それぞれのアンケート項目が推定出来ているかどうかの指標となる正解率はそれぞれの項目の最大値で「難しさ：0.41」「面白さ：0.55」「疲労：0.522」「集中度：0.505」であった。この実験結果を考慮すると「推定は出来ているが精度は低く、改善が必要」と結論せざるを得ない。

「推定が出来ている」とした理由は、第 4 章の調査で収集したアンケートは各問題に対しての難しさ・面白さ・疲労・集中度に対して「5 段階」で評価しており、もし学習中の学習者の表情とモチベーションに影響すると思われる上記の 4 要素に関連性が無く推定が不可能な場合、実質的に無作為な組み合わせとなり正解率は 0.2 前後になると考えられる為である。 [19]

しかし、完全に無作為・無関係な場合と比べて実験により得られた正解率が高いと入っても最高で 0.5 程度であるなら半分は不正解であり実用に耐えるものではない。したがって改善が必要だが、その改善にあたって解決する必要のあると思われる課題を次節に記した。

第7.2節 今後の課題

第7.2.1項 実験段階での課題

データの量の問題

機械学習の分野、特にニューラルネットワーク関連の研究や実用においては使用するデータは何万個単位である場合が多い。本研究において第 4 章で入手したデータは約 30 万個であるが第 5.2 節によって 453 個まで圧縮されてデータの個数が減少している。しかし圧縮自体はデータをニューラルネットワークに学習させるために必要なものであった以上、それ自体が問題ではなく、単にデータの収集数が少なかったことが適切な推定が出来なかった原因である可能性がある。

この課題の解決策としては調査の内容を専用の web アプリなどを開発して自動化し、インターネットを通じて協力者を募るなどして幅広くデータを集めることなどが考えられる。ただし後述のデータの質の問題（実験環境の隔離）と引き換えになることに留意しなければならない。

調査環境の隔離

第 4 章の調査において、JAIST の学生と外部の心理学研究機関の関係者に調査へ協力してもらったが、このとき外部機関の関係者に調査を行った際は、当該機関の施設内で調査を行った。このときは機関側の配慮で調査協力者の周辺はカーテンで囲われ、人の出入りや雑音が制限されるなど調査協力者にとって集中しやすい環境であった。

対して、JAIST の学生に調査への協力をしてもらった際には JAIST 内部のリフレッシュルームを用いていたため、ガラス越しに人の移動が見えたりした際に目線で追ってしまい、それが表情のデータにおける目線の記録としてノイズになるなどしてデータの質を引き下げた結果、適切な推定が出来なかった可能性がある。

この課題の解決策としてあらかじめ会議室を予約するなどの、環境の隔離が考えられる。ただしそれだけ調査にかかる労力や時間が増すことになり、前述のデータの量の問題と引き換えになることに留意しなければならない。

アンケート基準・定義の厳密化

第 3.4 節においてアンケート内容の基準の問題が発覚し、第 4 章の本調査において改善したが、それでもなおアンケート項目の定義や基準に対する質問が相次いだ事から改善が不十分であったと思われる。特に問題と思われるケースとして、アンケートにおける難しさの項目は調査協力者の主観での難しさを表現するものであるが、ある協力者は回答開始前に見せる低難易度の例題（第 4.2 節、データ収集段階の手続き 4）を基準として難易度の判断を行ったため、アンケートにおける難易度の判定は 1 から順につけられていたが、アンケート終了後の聞き取りにおいて協力者の主観では難易度を高く感じていたというケースがあった。

このような自体を防ぐためには、アンケートの項目について、用語の定義を更に厳密化することや、実際にアンケートに回答する際の例示を準備するなどして回答基準の個人差をできるだけ縮小することが考えられる。

虚偽記憶・記憶誤差の問題

本研究に協力してくれた外部の心理学研究機関からの指摘で、アンケートに回答された心理状態と実際の回答中の心理状態に差がある可能性が指摘された。

具体的には第 4.2 節の手順では CAB 問題回答後にそれを思い出しながらアンケートを書くこととなっているが、この際、思い出す為の補助として見せていた回答や解説が無意識下で記憶に影響を与え、記憶が変質した可能性があるとの指摘を受けた。例として、調査協力者が回答中にある問題を難しいと感じたが、アンケート回答中に解説を見たことでアハ体験を起こし、この結果難易度の低い問題と誤認してしまい、これが無意識下で行われるため後の聞き取りでも発覚しないというものである。

機関側の提案では調査の方法をアンケートではなく聞き取りにし、調査協力者が中立的な聞き取り調査の手法について技能を持つことでこの問題を解決できるとしているが、これは調査にかかる労力との兼ね合いを考慮する必要がある。

疑似相関の問題

第 4 章の調査によって、「難しさ」「面白さ」「疲労」「集中度」の 4 つについて、データを収集し、第 5 章にてこれらの推定を行ったが、このとき「難しさ」と「集中度」の項目について、ニューラルネットワークはそれ自体の推定を行っていない可能性が考えられた。原因に第 4 章の調査において出題される CAB の問題は、難易度の低い方から先に出題され、高いものほど後から出題されるような順番になっていたため、アンケートの難易度項目の判定が問題番号と比例するようになっており、これでは時間が進むに連れて判定が高くなっていく「疲労」と似たような判定の推移を持つことになり、ニューラルネットワークによる推測ではこの両者が混同された可能性がある。

これを防ぐためには CAB 問題の難易度と出題の順番を線形にしないといった方策が考えられるが、CAB 未経験者にとって問題への慣れを阻害する事になり、それが別の課題を引き起こす可能性もある。

第7.2.2項 データ分析段階での課題

機械学習アルゴリズムの最適化

第 5 章においてニューラルネットワークを用いてアンケート結果の推定を行い、低い正解率だったため、第 6 章において他の分析手法を提案したが、この内 RNN は本実験に付随して実際に実験を行っていた。しかし一般的な NN と比べて解析に極めて時間がかかるため仮に予測ができたとしても実用的ではない可能性がある。

この原因として RNN は隠れ層に過去の入力データを保持する関係上、専有するメモリ容量が大きく、データ量としても３０万個ちかいデータを読み込ませ計算させるため非常に時間がかかる（JAIST の計算サーバーを用いてもニューラルネットワークの学習５０回が 7 時間程度で終了するのに対し、RNN では 1 回の学習に 7 ～ 9 時間程度かかった）。学習回数を少なくすることで時間は短縮できるが、精度と引き換えになる上に、テストデータごとのばらつきも大きくなり推定結果に対して妥当性を保証できない。

原因としては RNN の性質そのものとして時間がかかりやすい事が挙げられるが何らかのプーリングやデータ量の削減をするなどして RNN の（時間に対して）効率的な方法を開発する必要がある。

謝辞

この場を借りて以下の方々にお礼を申し上げます。

長谷川忍准教授には本研究の主指導教員として研究に関する助言・関連研究のサーベイ・学内手続きに関する解説等、不肖を最後まで支えて頂きお礼申し上げます。

また、小谷一孔教授、及び白井清昭准教授には不肖では思い至らなかった課題の指摘を賜り、お礼を申し上げます。

また、調査に協力してくれた何君、藤本君に改めて感謝いたします。

また、今回調査に協力の上、重要な指摘をしてくださった心理療法研究所トマニの関係者各位にお礼申し上げます。

参照文献

[1] ウィキメディア財団, “コンピュータ支援教育,” 13 04 2017. [オンライン]. Available: https://ja.wikipedia.org/wiki/%E3%82%B3%E3%83%B3%E3%83%94%E3%83%A5%E3%83%BC%E3%82%BF%E6%94%AF%E6%8F%B4%E6%95%99%E8%82%B2. [アクセス日: 05 02 2019].

[2] ウィキメディア財団, “e ラーニング,” 29 01 2019. [オンライン]. Available: https://ja.wikipedia.org/wiki/E%E3%83%A9%E3%83%BC%E3%83%8B%E3%83%B3%E3%82%B0. [アクセス日: 05 02 2019].

[3] “Anki - Powerful, intelligent flash cards,” 17 10 2018. [オンライン]. Available: https://apps.ankiweb.net/. [アクセス日: 05 02 2019].

[4] duolingo, “Duolingo：英語、スペイン語、中国語など、いろんな言語を無料で学ぼう,” 11 2017. [オンライン]. Available: https://ja.duolingo.com/. [アクセス日: 06 02 2019].

[5] POLYGLOTS, Inc., “POLYGLOTS | ポリグロッツ |,” 2018. [オンライン]. Available: https://www.polyglots.net/. [アクセス日: 06 02 2019].

[6] 和田公人, 失敗から学ぶ e ラーニング, オーム社, 2004, pp. 46-62.

[7] 角. 考. 正. 導. 中村 和晃, “e-learning における学習者の動作観測に基づく主観的難易度の推定,” 電子情報通信学会論文誌. D, 情報・システム = The IEICE transactions on information and systems (Japanese edition) 93(5), 2010.

[8] 小. 一. 松. 辰. 堀口 祐樹, “e-learning における学習時の潜在的な意識変化の抽出システム,” 2008 年度人工知能学会全国大会（第 22 回）, 2008.

[9] 熊. 忠. 三. 正. 李 凱, “モーションセンサを用いた学習活動の状態推定手法の開発,” 教育システム情報学会誌 33 巻 2 号, 2016.

[10] 矢島邦昭, “E ラーニングのためのバイオセンシング研究,” e ラーニング高等教育連携(eHelp)全体会議, 2018.

[11] Megvii., “Face++ Cognitive Services,” 2019. [オンライン]. Available: https://www.faceplusplus.com/. [アクセス日: 06 02 2019].

[12] SPI ノートの会, '20　必勝・就職試験！CAB・GAB 完全突破法！, 洋泉社, 2018.

[13] 巣籠悠輔, 詳解ディープラーニング　TensorFlow・Keras による時系列データ処理, マイナビ出版, 2017.

[14] id:nkdkccmbr, “ニューラルネットワーク・DeepLearning などの画像素材,” 06 10 2016. [オンライン]. Available: http://nkdkccmbr.hateblo.jp/entry/2016/10/06/222245#%E5%9F%BA%E6

%9C%AC%E7%9A%84%E3%81%AA%E3%83%8B%E3%83%A5%E3%83%BC%E3%83%A9%E3%83%AB%E3%83%8D%E3%83%83%E3%83%88%E3%83%AF%E3%83%BC%E3%82%AF. [アクセス日: 07 02 2019].

[15] 大人になってからの再学習, “Google の猫認識 (Deep Learning),” はてなダイアリー, 08 06 2013. [オンライン]. Available: http://zellij.hatenablog.com/entry/20130608/p1. [アクセス日: 07 02 2019].

[16] Google, “TensorFlow の概要,” 2018. [オンライン]. Available: https://www.tensorflow.org/. [アクセス日: 06 02 2019].

[17] Keras, “Keras Documentation,” MkDocs, 2019. [オンライン]. Available: https://keras.io/ja/. [アクセス日: 06 02 2019].

[18] Amazon, “モデルフィット: アンダーフィットとオーバーフィット,” 2019. [オンライン]. Available: https://docs.aws.amazon.com/ja_jp/machine-learning/latest/dg/model-fit-underfitting-vs-overfitting.html. [アクセス日: 07 02 2019].

[19] rikunora, “機械学習.vs.乱数,” 06 01 2016. [オンライン]. Available: http://d.hatena.ne.jp/rikunora/20160106/p1. [アクセス日: 07 02 2019].

[20] 萩. 将. 齋藤 雅裕, “類推を行う言語処理ニューラルネットワーク,” 電子情報通信学会技術研究報告. NC, ニューロコンピューティング 108(480), 1-6, 2009-03-04, 2009.

[21] 杉. 村松辰真, “顔情報を利用した集中度測定システムの効果検証,” 第 74 回全国大会講演論文集 2012(1), 2012.

[22] 菅沼義昇, “ニューラルネットワーク,” 2016. [オンライン]. Available: https://www.sist.ac.jp/~suganuma/kougi/other_lecture/SE/net/net.htm. [アクセス日: 07 02 2019].

[23] D. R.Ramya, “ Student Engagement Identification based on Facial Expression Analysis using 3D Video/Image of students,” TAGA JOURNAL Vol.14, 2018.

[24] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster , J. R. Movellan, “The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions,” IEEE Transactions on Affective Computing vol.5, 2014.

付録 A　収集されたデータ

A1 予備調査で使用したアンケート

氏名＿＿＿＿＿＿＿＿＿＿＿＿　　日付　　年　　月　　日

裏面の例を参考にして、記入してください。

	難易度					面白さ					疲労				
	簡単	やや簡単	普通	やや難しい	難しい	つまらない	ややつまらない	普通	やや面白い	面白い	元気	やや元気	普通	やや疲れた	疲れた
1 問目															
2 問目															
3 問目															
4 問目															
5 問目															
6 問目															
7 問目															
8 問目															
9 問目															
10 問目															
11 問目															
12 問目															
13 問目															
14 問目															
15 問目															
16 問目															
17 問目															
18 問目															
19 問目															
20 問目															
21 問目															
22 問目															
23 問目															
24 問目															
25 問目															
26 問目															
27 問目															
28 問目															
29 問目															
30 問目															

表面

「E ラーニング学習者における表情由来の心理情報抽出に関する検討」の実験に関するアンケート

表面には、実験で解いた問題それぞれについて
「どれくらい難しかったか」
「どれくらい面白かったか」
「どれくらい疲れていたか（問題を解き終わった時点で)」
について、５段階で回答してもらいます。
以下の例のように当てはまるものに「✓」をつけてください。
また、問題が全て解いていない場合、解いていない問題の番号に以下の例のように斜線を引いてください。

例：
「難しさ」は全体的に難しいが２８問目から更に難しくなった
「面白さ」は全体的に普通だが、２７問目だけはつまらなかった
「疲れたか」について最初は普通だったが２７問目から疲れ始め、２９問でピークに達した。
時間が足りなくて３０問目は解くことが出来なかった。

	難易度					面白さ					疲労				
	簡単	やや簡単	普通	やや難しい	難しい	つまらない	ややつまらない	普通	やや面白い	面白い	元気	やや元気	普通	やや疲れた	疲れた
26 問目				✓				✓					✓		
27 問目				✓		✓								✓	
28 問目					✓			✓						✓	
29 問目					✓			✓							✓
~~30 問目~~															

裏面

A2 本調査で使用したアンケート

氏名＿＿＿＿＿＿＿＿＿＿＿＿　　日付　　年　　月　　日

裏面の例を参考にして、記入してください。

	難易度					面白さ					疲労					集中				
	簡単	やや簡単	普通	やや難しい	難しい	つまらない	ややつまらない	普通	やや面白い	面白い	元気	やや元気	普通	やや疲れた	疲れた	集中できた	やや集中できた	普通	やや集中できない	集中できない
1																				
2																				
3																				
4																				
5																				
6																				
7																				
8																				
9																				
10																				
11																				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19																				
20																				
21																				
22																				
23																				
24																				
25																				
26																				
27																				
28																				
29																				
30																				

「E ラーニング学習者における表情由来の心理情報抽出に関する検討」の実験に関するアンケート

表面には、実験で解いた問題それぞれについて
「どれくらい難しかったか」
「どれくらい面白かったか」
「どれくらい疲れていたか（問題を解き終わった時点で）」
について、５段階で回答してもらいます。
以下の例のように当てはまるものに「✓」をつけてください。
また、問題が全て解いていない場合、解いていない問題の番号に以下の例のように消し線を引いてください。

	難易度					面白さ					疲労					集中				
	簡単	やや簡単	普通	やや難しい	難しい	つまらない	ややつまらない	普通	やや面白い	面白い	元気	やや元気	普通	やや疲れた	疲れた	集中できた	やや集中できた	普通	やや集中できない	集中できない
25		✓								✓			✓				✓			
26		✓				✓								✓				✓		
27			✓				✓							✓					✓	
28			✓				✓								✓		✓			
29				✓				✓							✓			✓		
~~30—~~																				

付録 B　追試験及び新たな課題

第B.1節 概要

本付録は本文執筆後に行った追試験及び新たに発見された課題に関して述べるものである。

第B.2節 ニューラルネットワークの追試験

本節では第 4 章にて入手したデータを第 5.5 節の本実験と同様の条件でアンケート結果の推定を行った結果を示す。この条件を下記に示す。

- 入力層のノードは 83 個（入力データの次元数と同じ）
- 隠れ層はシグモイド関数を用いノードは 83 個（入力データの次元数と同じ）
- 出力層はソフトマックス関数を用い、ノードは 5 個
- 学習率は 0.2
- 隠れ層の層数を 0 ～ 5 層に変更し、比較する。
- 全データの中から３０個をテストデータとして抜き出し、残りを訓練データとしてニューラルネットの学習に用いる。
- 抜き出すテストデータは５０回変更し、それぞれのテストデータごとの正解率を取り、その平均値を見る。

また、本節の実験では平均正解率に加え、各テストデータを用いたアンケート結果の推定において、調査協力者が記入したアンケート結果（正答データ）とニューラルネットにより推定されたアンケート結果（予測データ）の組み合わせを収集し、これを比較した。

第B.2.1項　　「難しさ」の項目における追試験結果

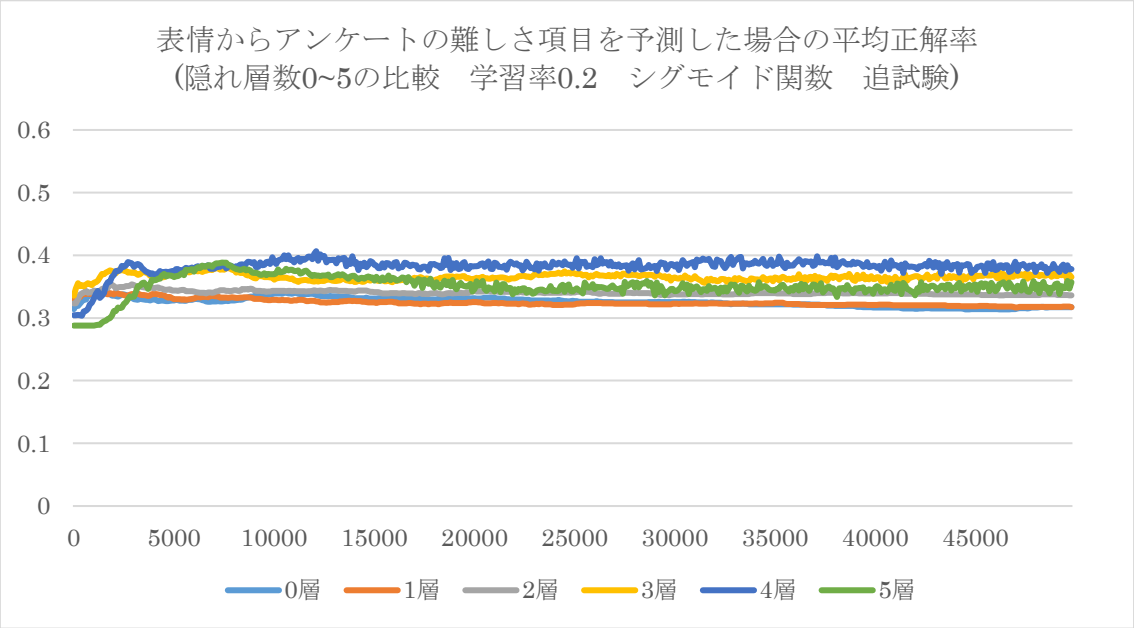

図 0-1 「難しさ」の項目追試験における平均正解率

第 5.5.1 項と比較すると図 5-10 では隠れ 2 層が最も正解率が高かったのに対し、図 0-1 では 4 層が最も正解率が高くなっている。また、正解率の全体的な範囲は図 5-10 と同じく 0.3~0.4 の範囲に収まっているが、エポック数を重ねたあとの過学習の影響があまり無い事や、エポック数 35000~45000 付近において各層の正解率が収束していない等の差が見られる。

原因としてデータの抜き出しがランダムであるため、それによって差が生じた可能性はあるが、その場合いかなるデータの選抜方法が適正であるかという課題が生じる。

また、本項において最も高い平均正解率を記録した 4 層について、テストデータ５０回分の正解・不正解の内訳を分析したグラフを以下に示す。

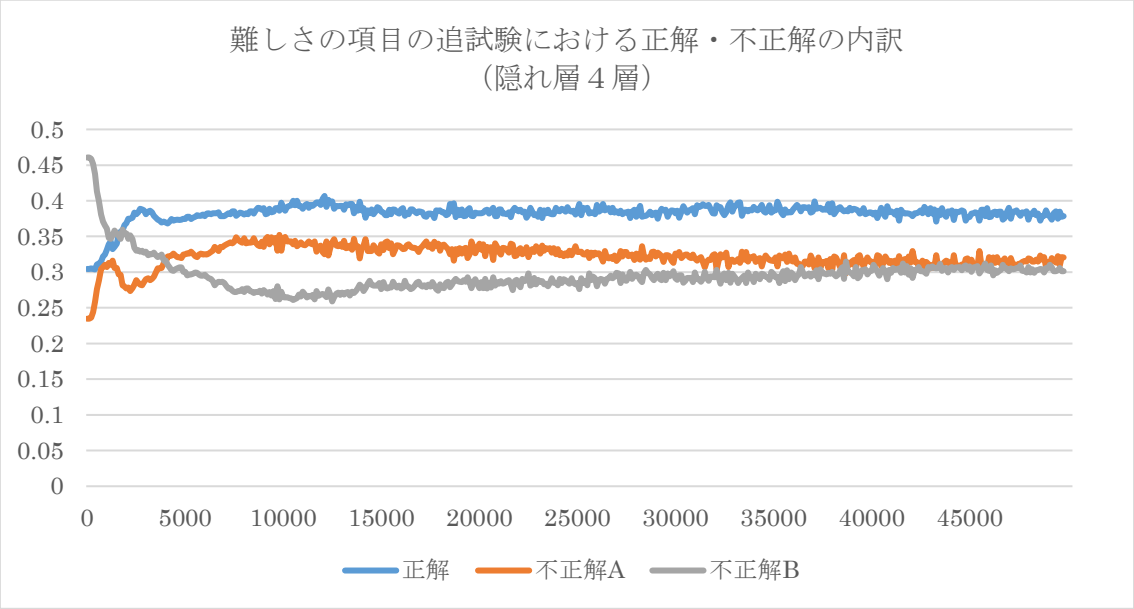

図 0-2 「難しさ」項目４層追試験時の正解・不正解の内訳

図 0-2 において、テストデータ５０回分の正解・不正解のうち、不正解を「予想がアンケートと１だけずれている」ものを不正解 A、「予想がアンケートと２以上ずれている」ものを不正解 B とし、全体に占める割合を計算した。このグラフから、全体として正解と不正解 A は序盤から上昇し、不正解がそれに対となるように下落する様子が確認できる。

第B.2.2項　　「面白さ」の項目における追試験結果

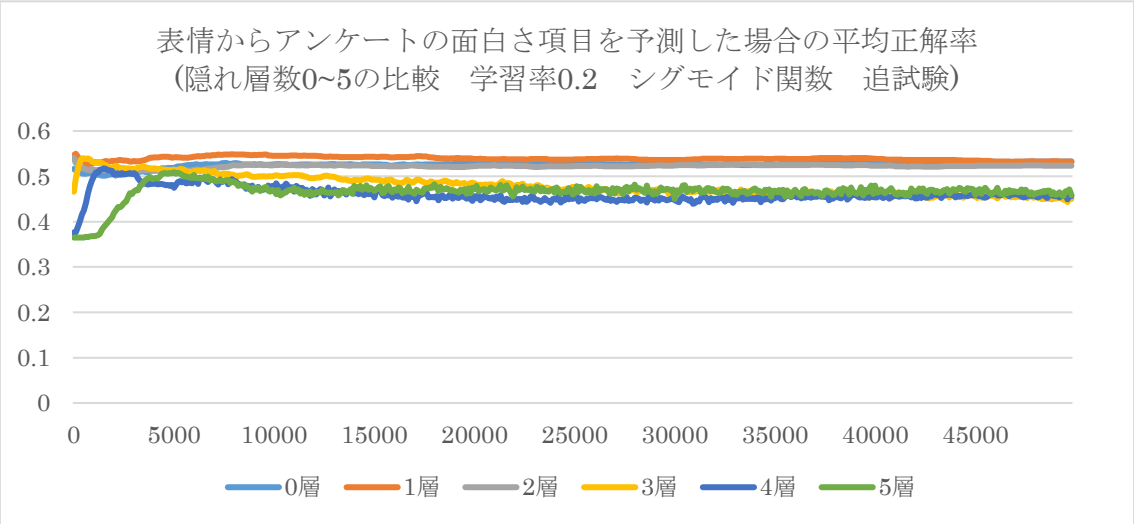

図 0-3 「面白さ」の項目追試験における平均正解率

第 5.5.2 項と比較すると、図 5-11 では平均正解率の最高を記録したのは隠れ層 0 層の場合のみで、それ以外は過学習の影響を受けて徐々に平均正解率が下落していたが、本項では隠れ層 0 層は平均正解率が低く、またエポック数を重ねるごとに低下しているが、1～2 層は図 5-11 の 0 層のように、エポック数を重ねても正解率の低下が見られない。3～5 層は後半で低下しており、この傾向は図 5-11 と共通している。ただし 5 層に関しては図 5-11 では前半で最高値を記録しない（エポック数を重ねるごとに比例的に平均正解率が上がっている）という相違がある。

また、本項において最も高い平均正解率を記録した 1 層について、テストデータ 5 0 回分の正解・不正解の内訳を分析したグラフを以下に示す。

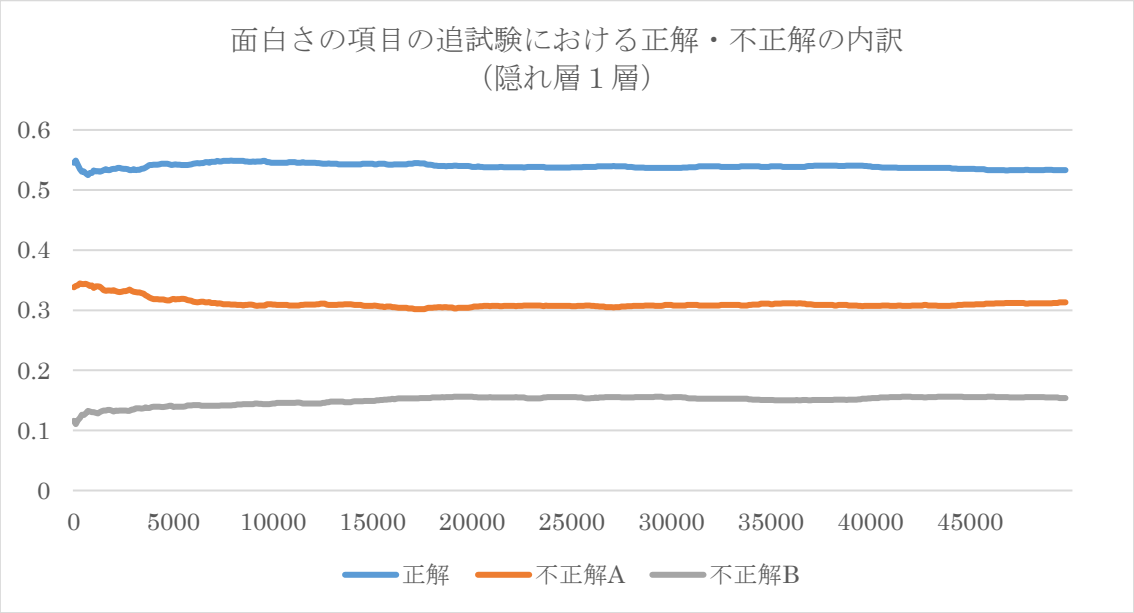

図 0-4 「面白さ」項目 1 層追試験時の正解・不正解の内訳

図 0-4 の凡例（正解・不正解 A・不正解 B）の基準は図 0-2 と同じである。図 0-4 を見ると、図 0-2 と同じように、全体的な内訳の割合は正解＞不正解 A＞不正解 B となっているが、図 B 2 と違い値の変動がほとんど無く、特に不正解 B が最初から最低となっている点に相違がある。しかしそれでも正解率が最高でも 0.54 と実用的な高さではない。この詳細を分析するために、正答データと予想データの組み合わせを一部抜粋した表を以下に示す。

表 0-1 「面白さ」項目追試験 1 層 5000 エポックにおける予測・正答の組み合わせ

エポック数 5000

予想＼正答	1	2	3	4	5	sum
1	7	9	8	17	3	44
2	4	120	45	53	17	239
3	6	11	360	135	21	533
4	4	21	116	198	64	403
5	6	12	36	76	99	229
sum	27	173	565	479	204	1448

表 0-1 は正解を青色、不正解 A を黄色、不正解 B を赤色に着色している。表 0-1 を見ると、正答（実際のアンケート回答）が 1 の組み合わせの数が少なく、予想（ニューラルネットによるアンケート回答の予想）が 1 の組み合わせも少ない事がわかる。これはニューラルネット側で予想が 1 の組み合わせを出力するための材料が少ないことから予想 1 の組み合わせの総数自体が減っていると考えられる。それにより結果的に正答が 3・4・5 の組み合わせにおいて不正解 B となる予想 1 の組み合わせの数が減少し、結果的に不正解 B の割合が低くなっていると推測される。

この現象はニューラルネットが正しく正答 1 を予想しようとしているとは見なせないが、これは抜き出したテストデータが偏ったのではなく第 4 章の本調査で入手したアンケートデータの「面白さ」の項目において、元々 1 と記入されたデータが非常に少ないことに由来する。

対策として、テストデータにおいて正答の数を均等にすることなどが考えられるが、その評価が正答であるかどうか等に考察を要する。

第B.2.3項　　「疲労」の項目における追試験結果

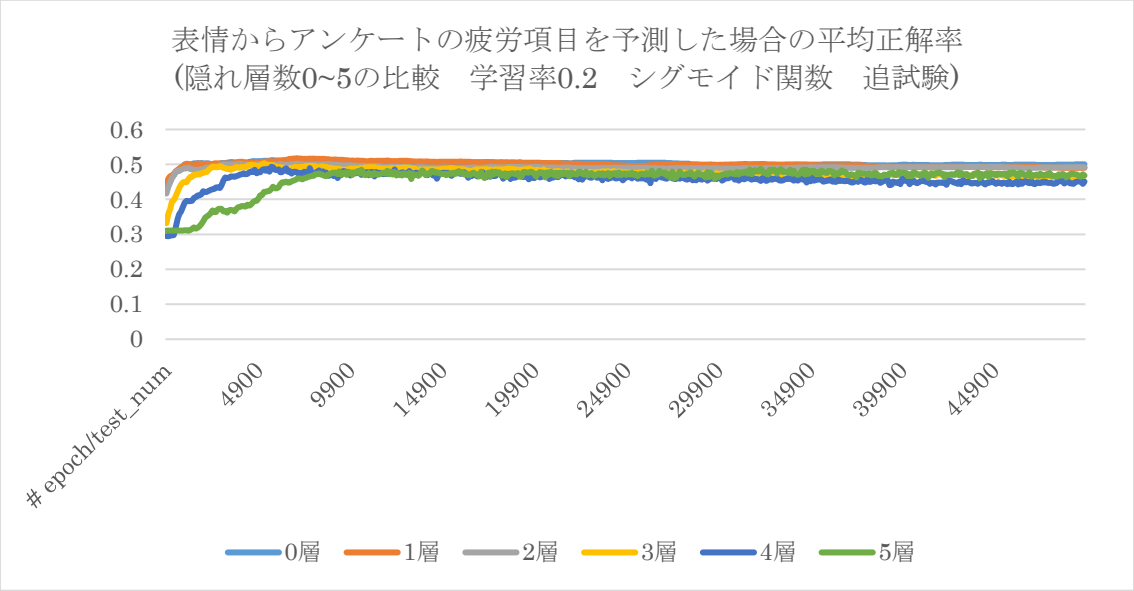

図 0-5 「疲労」の項目追試験における平均正解率

第 5.5.3 項と比較すると、図 5-12 では隠れ層 2 層が最高の平均正解率を記録しているが、図 0-5 では（グラフでは分かりづらいが）0 層が 6900 エポックで平均正解率最高値 0.512 を記録している。しかし 0 層～2 層がエポック数を重ねても 3～5 層ほど平均正解率が低下しない点で図 5-12 と図 0-5 は共通している。また、以下に表に示す最高平均正解率の比較において、0 層～2 層の平均正解率の差が少ない事などから、0 層～2 層において生じる隠れ層と平均正解率の差は誤差であると考えられる。

表 0-2 「疲労」項目の本実験と追試験における最高平均正解率の比較

隠れ層＼実験	第 B.2 節　追試験	第 5.5 節　本実験
0 層	0.512445 (6900 エポック)	0.502676 (5400 エポック)
1 層	0.517671 (7100 エポック)	0.510131 (12400 エポック)
2 層	0.501944 (4000 エポック)	0.522046 (5200 エポック)
3 層	0.503620 (5400 エポック)	0.491092 (22900 エポック)
4 層	0.493324 (5700 エポック)	0.479062 (25700 エポック)
5 層	0.488104 (32300 エポック)	0.464031 (49900 エポック)

また、本項において最も高い平均正解率を記録した０層と最低を記録した４層について、テストデータ５０回分の正解・不正解の内訳を分析したグラフを以下に示す。

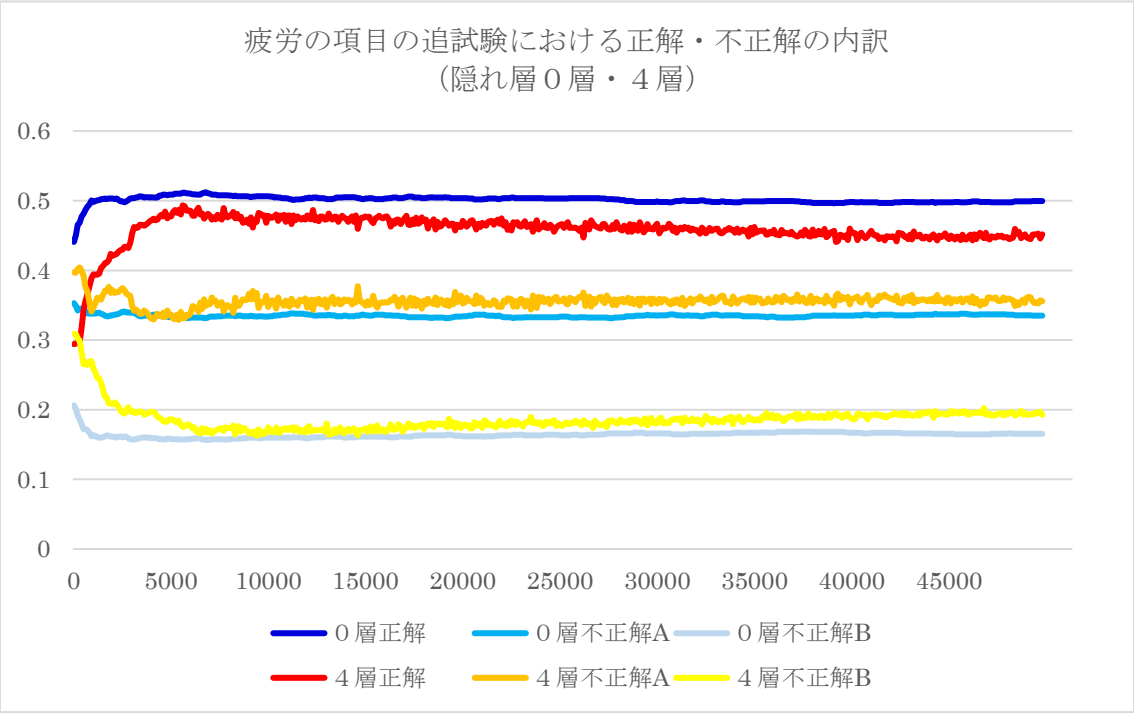

図 0-6 「疲労」項目１層・４層追試験時の正解・不正解の内訳

図 0-6 を見ると、０層では各割合が正解＞不正解 A＞不正解 B に並び一度も交差しない事が見えるが、４層では図 0-2 のように序盤で交差している。ただし図 0-2 の序盤では不正解 B が最高であったのに対し、図 0-6 では不正解 A が最高であり、この割合が大きく変動することはなかった。

考察としては、与えられたテストデータに対して最序盤の学習と予想結果において正解の割合が高いものはその状態が維持される、ある種の局所解に陥っている可能性がある。この場合、局所解に嵌まる原因として初期値（ランダムに選抜された訓練データとテストデータ）が考えられるが、選抜方法をいかなる方式に変更するか、及びその正当性と妥当性をいかに保証するかが新たな課題となる。

第B.2.4項　　「集中度」の項目における追試験結果

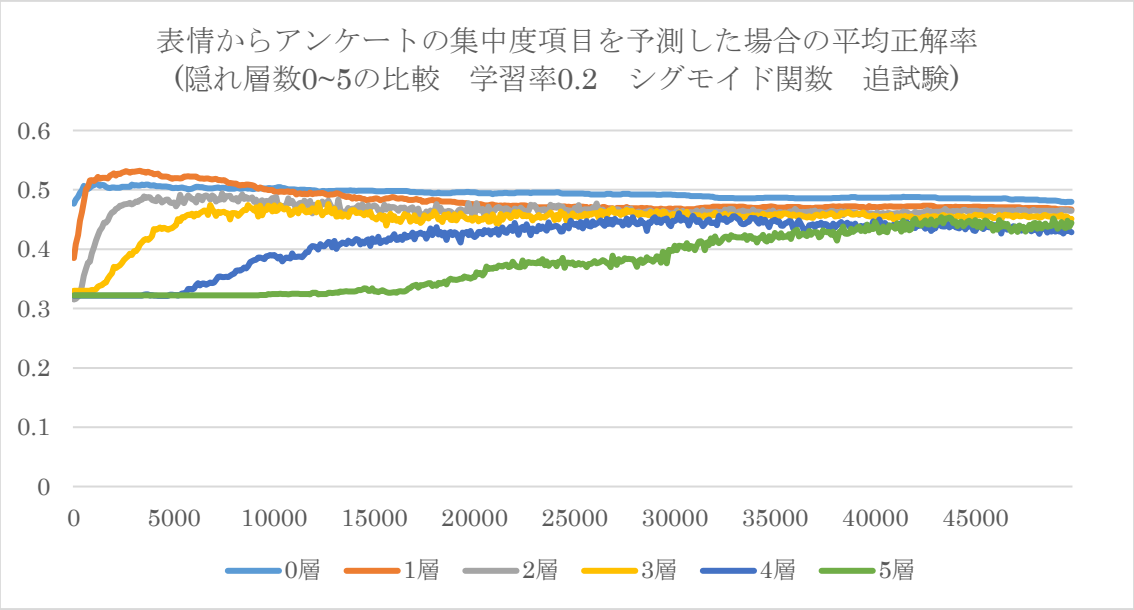

図 0-7 「集中度」の項目追試験における平均正解率

第 5.5.4 項と比較すると、図 5-13 よりも図 0-6 のほうが隠れ層 1 層の最高平均正解率が高い等の差があるが、両図とも 0 層の平均正解率が下がりにくい事や、1 層の平均正解率は一時的に最高値になったあと 0 層以下に収まること、5 層が全体的に低い等の共通点が見られる。第 B.2.1 項～第 B.2.2 項では本実験と追試験で平均正解率の傾向に差があったのに対し、こちらは傾向にあまり差が見られないことから、比較的予測が安定していると思われる。

本項において最も高い平均正解率を記録した 1 層について、テストデータ 50 回分の正解・不正解の内訳を分析したグラフを以下に示す。

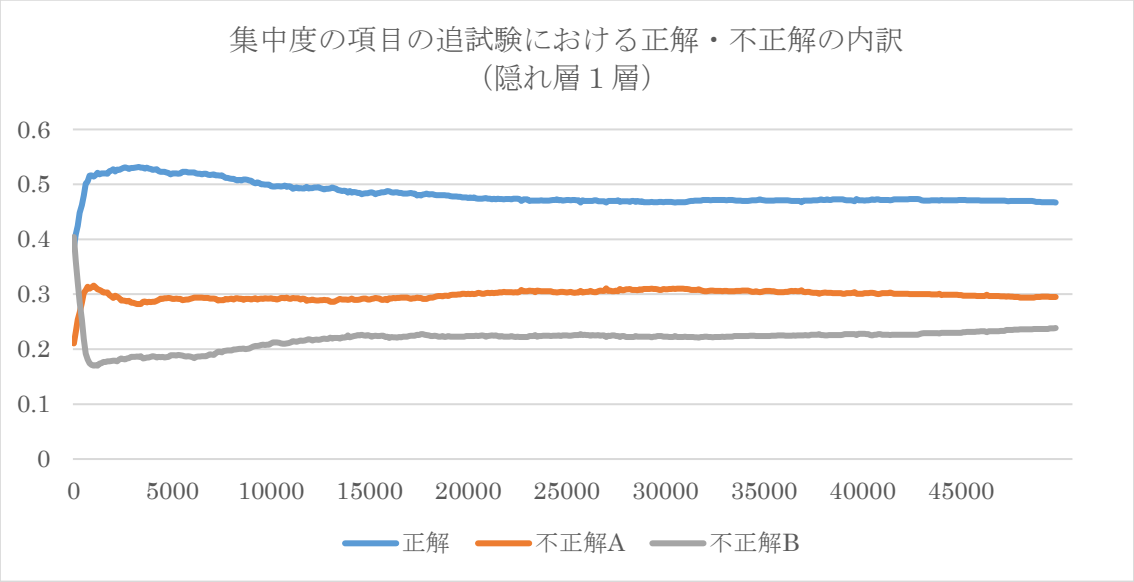

図 0-8 「集中度」項目１層追試験時の正解・不正解の内訳

図 0-8 を見ると、最序盤では不正解 B の割合が最も多く、直後に不正解 B と正解の割合が交差するという図 0-2 と似た特徴を有している。ただしこちらのほうの最初の不正解 B と正解の割合の差は大きくない。また図 0-4・図 0-6 において不正解 B の割合は概ね 0.1~0.2 の範囲に収まっていたが、図 0-8 では 0.2 を超えており、不正解 A と不正解 B の割合の差が比較的小さいなどの差がある。

第B.2.5項　SVM・KNN 法との比較

本項では第 B.2.1 項～第 B.2.4 項に記載したニューラルネットワークの追試験結果における平均正解率の最高値と、SVM（サポートベクターマシン）と KNN 法（k 近傍法）を用いて表情データからアンケートデータを推定した際の正解率を比較する。この 2 法についても、テストデータの抜き出しをランダムに 50 回行い、各テストデータの正解率の平均値を比較する。また、KNN 法では K 値（近傍個数）を１～１９９まで変化させ、最も高い平均正解率を比較の対象とする。この時の K 値ごとの平均正解率の比較を、縦軸を正解率・横軸を K 値とした以下の図に示す。

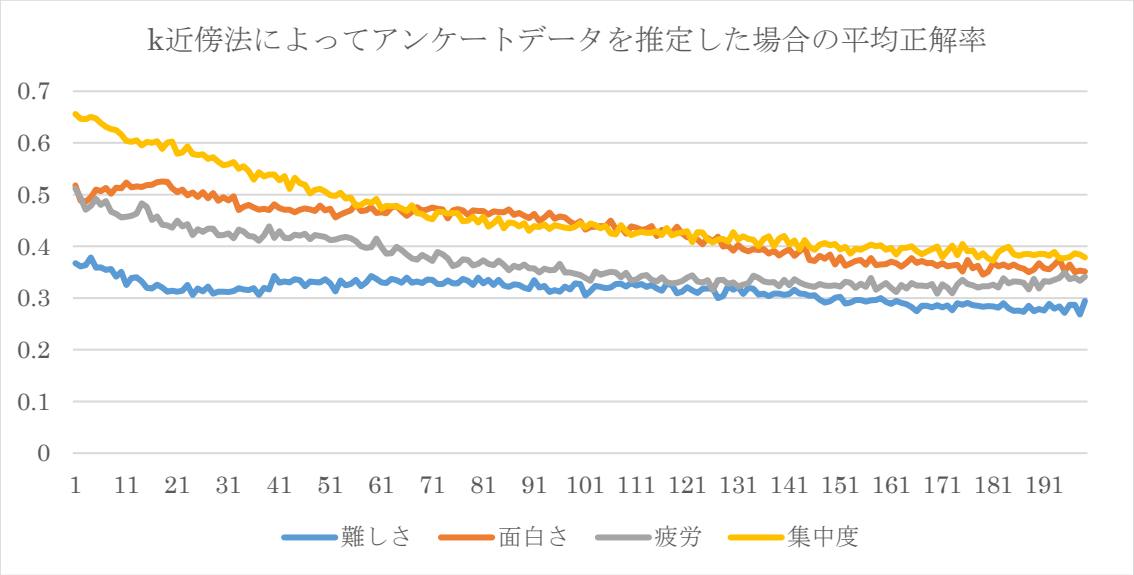

図 0-9　k 近傍法を用いた推定の平均正解率

上記の KNN 法の図と SVM の平均正解率の結果を比較した表を以下に示す。

表 0-3　各手法の平均正解率比較

手法＼推定項目	難しさ	面白さ	疲労	集中度
NN 本実験	0.419169 (隠れ 2 層)	0.551034 (隠れ 0 層)	0.522046 (隠れ 2 層)	0.50563 (隠れ 0 層)
NN 追試験	0.406618 (隠れ 4 層) (12200 ｴﾎﾟｯｸ)	0.549803 (隠れ 1 層) (200 ｴﾎﾟｯｸ)	0.51767 (隠れ 1 層) (7100 ｴﾎﾟｯｸ)	0.53220 (隠れ 1 層) (3400 ｴﾎﾟｯｸ)
SVM	0.323516	0.504175	0.422197	0.523956
KNN 法	0.378461 (k=4)	0.525494 (k=18)	0.511868 (k=1)	0.656043 (k=1)

表 0-3 を見ると「集中度」以外は NN（ニューラルネット）が最も高い平均正解率を持っている用に見える。また、NN においては追試験における「難しさ」の予測以外すべて隠れ層 2 層以下という特徴を有する。

第B.3節 GRU による推定

本節では第 4 章にて入手したデータを第 6.2 節で述べた GRU の特性に基づき、第 5.2 節にて述べた圧縮処理を行わずに推定を行った。条件は以下の通りである。

- 入力層のノードは 31 個（入力データの加工前の次元数と同じ）
- 隠れ層はシグモイド関数を用い、ノードは 32,64,128 個の 3 通り
- 出力層はソフトマックス関数を用い、ノードは 5 個
- 学習率は 0.2
- 隠れ層の層数を 1 ～ 3 層に変更し、比較する。
- 全データの中から３０個をテストデータとして抜き出し、残りを訓練データとしてニューラルネットの学習に用いる。（この時用いるデータとは問題ごとの表情データの集まりであり、集まりごとの表情データ（フレーム）の個数は可変である。）
- 抜き出すテストデータは 5 回変更し、それぞれのテストデータごとの正解率を取り、その平均値を見る。

以上の条件において、各アンケート項目に対し隠れ層のノード数と隠れ層の層数ごとに平均正解率の最高値を記録した。

第B.3.1項　　「難しさ」の項目における試験結果

ノード数＼層数	１層	２層	３層
３２個	0.391111111	0.404444444	0.422222222
６４個	0.364444444	0.391111111	0.395555556
１２８個	0.368888889	0.368888889	0.4

第 B.2.1 項と比較すると、3 層 32 個において平均正解率の最高値がやや上回っている。また、全体的にノード数 128 個において正解率が芳しく無く、ノード数が少ないほうが良いように見える。

第B.3.2項　　「面白さ」の項目における試験結果

ノード数＼層数	１層	２層	３層
３２個	0.493333333	0.537777778	0.506666667
６４個	0.475555556	0.484444444	0.515555556
１２８個	0.475555556	0.524444444	0.511111111

第 B.2.2 項と比べると、2 層 32 個において第 B.2.2 項の最高平均正解率に近づいているが、それ以外は殆下回っている。また、第 B.3.1 項と比較した場合、1~2 層においてノード数が増えると平均正解率が下がる傾向が共通している。

第B.3.3項　　「疲労」の項目における試験結果

ノード数＼層数	１層	２層	３層
３２個	0.511111111	0.488888889	0.515555556
６４個	0.511111111	0.515555556	0.493333333
１２８個	0.533333333	0.497777778	0.474074074

第 B.2.3 項と比較すると、1 層 128 個において正解率がやや上回っている。また、第 B.3.1 項、第 B.3.2 項と比較すると、1 層 128 個とノード数が多い方の正解率が高い。しかし 2 層と 3 層では 128 個が最高ではない以上、誤差の可能性もある。

第B.3.4項　　「集中度」の項目における試験結果

ノード数＼層数	１層	２層	３層
３２個	0.293333333	0.342222222	0.275555556
６４個	0.346666667	0.288888889	0.275555556
１２８個	0.253333333	0.271111111	0.316666667

第 B.2.4 項と比較すると、平均正解率が半分程度しかなく、第 B.3.1 項～第 B.3.3 項のいずれの結果と比較しても突出して低い。最高値は 1 層 64 個であるが、3 層において 128 個が 3 層内で最高値を記録するなど、第 B.3.1 項～第 B.3.3

項で見られたノード数が増えるにつれて平均正解率が下がる傾向から逸脱している。アンケート項目「集中度」に関しては第 B.2.5 項の比較において唯一正解率の比較が逆転するなど現象から、何らかの特異性を持っているものと思われる。

第B.3.5項　　総括

全体として、GRU を用いた結果、第 B.2 節の NN を用いた実験と比べ、それほど大きい改善は見られなかった。ただし、この実験では GRU の計算時間が長大であるという事情からテストデータの抜き出し回数が 5 回と少なく設定されている。後の第 B.5.1 項で述べる通り、テストデータの抜き出しが小回数のランダムである限り再現性が無いため、NN との比較優位を論ずるには早計であると考えられる。

第B.4節 非線形相関による分析

本節では表情データとアンケートデータの関連性を分析するために、MIC(Maximal information coefficient) [21]と呼ばれる指標を用いて非線形の相関係数を計算した。この計算結果を以下の表に示す。

表 0-4　表情の各パラメータとアンケート結果に対する MIC

			相関	相関	相関	相関
			d_難しさ	i_面白さ	t_疲労	c_集中度
emothion	sadness	avg	0.146929	0.169227	0.155943	0.181091
		var	0.160775	0.18936	0.163005	0.215444
		max	0.159422	0.188179	0.145371	0.20465
		min	0.070994	0.077669	0.041128	0.112026
	netural	avg	0.14852	0.171005	0.16946	0.188946
		var	0.19375	0.185722	0.161535	0.230658
		max	0.164072	0.162017	0.166713	0.248626
		min	0.123793	0.124377	0.104767	0.100881
	disgust	avg	0.159601	0.184456	0.183009	0.229085
		var	0.150076	0.190943	0.17256	0.17981
		max	0.156812	0.214911	0.165566	0.246317

		min	0.1576	0.225576	0.140688	0.194743
	anger	avg	0.196335	0.182442	0.22835	0.243578
		var	0.207886	0.173865	0.183471	0.183531
		max	0.15042	0.184025	0.158441	0.155762
		min	0.130395	0.157888	0.146203	0.223798
	surprise	avg	0.165021	0.207224	0.192589	0.212259
		var	0.186935	0.203712	0.158644	0.217526
		max	0.1783	0.19233	0.150153	0.181658
		min	0.062209	0.097609	0.040357	0.083774
	fear	avg	0.181774	0.160229	0.184704	0.16983
		var	0.19694	0.194065	0.17134	0.171457
		max	0.212914	0.201964	0.188057	0.15739
		min	0.070084	0.129993	0.066392	0.114473
	happiness	avg	0.17445	0.183854	0.206385	0.191237
		var	0.182546	0.189694	0.160119	0.201826
		max	0.134626	0.190812	0.148232	0.197165
		min	0.06136	0.137824	0.053529	0.095231
eyegaze	right_eye_x	avg	0.173512	0.239595	0.205836	0.206684
		var	0.215014	0.217702	0.208758	0.21787
		max	0.222804	0.182895	0.215027	0.226775
		min	0.180798	0.19962	0.198084	0.167973
	right_eye_y	avg	0.175615	0.289313	0.207508	0.30494
		var	0.17792	0.19028	0.175252	0.214562
		max	0.139747	0.15798	0.149509	0.202529
		min	0.197537	0.22634	0.192179	0.368358
	right_eye_z	avg	0.16537	0.318498	0.224668	0.317716
		var	0.244951	0.196934	0.18388	0.189357
		max	0.111008	0.164308	0.111795	0.210972
		min	0.169886	0.201877	0.184319	0.130878
	left_eye_x	avg	0.203769	0.305409	0.210481	0.214477
		var	0.17112	0.265819	0.187439	0.229458
		max	0.230601	0.20626	0.197251	0.185536
		min	0.175343	0.177968	0.190898	0.177681
	left_eye_y	avg	0.188409	0.244415	0.230163	0.185007
		var	0.162974	0.189234	0.153672	0.166331

		max	0.151146	0.143171	0.155248	0.207962
		min	0.186115	0.195878	0.207611	0.200699
	left_eye_z	avg	0.182257	0.328892	0.199645	0.194627
		var	0.184531	0.294234	0.189541	0.239222
		max	0.123023	0.15725	0.146632	0.119401
		min	0.231118	0.167303	0.190124	0.205902
retangle	width	avg	0.235632	0.277919	0.231938	0.314148
		var	0.141576	0.187557	0.174157	0.164683
		max	0.164962	0.207659	0.144205	0.215061
		min	0.13382	0.256345	0.180678	0.270422
	height	avg	0.235632	0.277919	0.231938	0.314148
		var	0.141576	0.187557	0.174157	0.164683
		max	0.164962	0.207659	0.144205	0.215061
		min	0.13382	0.256345	0.180678	0.270422
headpose	pitch	avg	0.169955	0.181631	0.24631	0.275847
		var	0.179558	0.211226	0.209878	0.174569
		max	0.168451	0.213182	0.188739	0.251575
		min	0.194563	0.225563	0.22307	0.29126
	roll	avg	0.164928	0.247174	0.23633	0.226208
		var	0.184146	0.18244	0.188003	0.20806
		max	0.164225	0.23921	0.201201	0.21744
		min	0.162077	0.203759	0.213264	0.182206
	yaw	avg	0.179895	0.239099	0.203172	0.231408
		var	0.150784	0.178148	0.157114	0.181964
		max	0.152818	0.193553	0.203812	0.219153
		min	0.159991	0.294966	0.201508	0.169788
mouth	close	avg	0.164281	0.18629	0.208792	0.19722
		var	0.164016	0.177255	0.186413	0.17821
		max	0.105429	0.128155	0.098075	0.199266
		min	0.147855	0.209483	0.165077	0.192086
	oppen	avg	0.180006	0.165732	0.156645	0.234876
		var	0.18204	0.1784	0.163509	0.159876
		max	0.1435	0.142057	0.185459	0.144618
		min	0.082468	0.101291	0.073982	0.113751
	other	avg	0.165194	0.172684	0.177908	0.209528

		var	0.135916	0.152703	0.18107	0.25649
		max	0.131254	0.193059	0.206256	0.23384
		min	0.036517	0.046439	0.041497	0.068499
age	age	age	0.127338	0.270842	0.315696	0.242545
gender	gender	gender	0.041354	0.106613	0.032891	0.089965
glass	glass	glass	0.09509	0.12786	0.106079	0.027587

表 0-4 を見るあたり、すべてのアンケート項目において、目線（eyegaze）の項目において弱い相関係数が確認されていること、及び表情の動的変化と直接関係のない年齢（age）の項目がアンケートの疲労の項目で弱い相関が確認されている点が特徴的である。このことから、より精度の高い予測を行う事や、表情と学習心理の関連性を詳細にするにあたって目線のデータを重視すべきであると考えられる。

また、表情の無表情（natural）と怒り（anger）が集中度と弱い相関が確認されており、これは集中している時は感情変化が少ない、もしくは集中している時のいわゆる真面目な顔が怒っている顔と類似している可能性があると考えられる。しかし本研究で用いた Face++は表情を 7 つの成分に分解した結果しか得られないため、次回ではその表情のもととなる表情筋の分割線などの、より詳細な表情データを用いることで高精度な予測ができる可能性があると考えられる。

第B.5節 新たな課題

第B.5.1項　訓練データ・テストデータ分別方法の妥当性

第 B.2.1 項～第 B.2.4 項の追試験では、第 5.5.1 項～第 5.5.4 項の本実験とほぼ同様の条件で実験を行ったにもかかわらず隠れ層ごとの平均正解率の順番が違ったり、平均正解率の推移傾向が違ったり等、再現性が確認できなかった。

この原因として、テストデータの選抜がランダムであるため、たまたま都合の良い・悪いデータが抜き出された結果、条件や手法とは無関係に平均正解率の増減が生じた可能性がある。このため、機械学習に用いる訓練データとテストデータのランダムな選抜が大数の法則が成立する程度の十分な回数の選抜を行うか、別の定性的な選抜方法を開発する必要がある。

提案としては、すべてのデータの内、組み合わせとなる個数分だけテストデータとして抜き出し、それ以外をすべて訓練データとして扱い、すべてのテストデ

ータが定数回だけ使用される方法が考えられる。例えば難易度の項目を予想する場合、全 453 データに対して、「テストデータ１：訓練データ４５２」という分け方で、すべてのデータが 1 回だけテストデータとして使用されるような場合、試行回数は４５３回である。これが「テストデータ２：訓練データ４５１」でも１０２３７８回である。ただしこの方法が妥当であるかどうかは別に議論を要する。

第B.5.2項　　連続値出力の方法とその評価方法の妥当性

第 B.2 節の追試験において正解率だけでなく正答と予測の組み合わせのデータも入手した結果、表情からアンケート結果を正しく推定できなかったデータであってもその誤差範囲が 1 である「惜しい」データが多く、完全に予測ができていないわけではないことが明らかになった。

また、本研究ではニューラルネットを用いたアンケートデータの予想において、ニューラルネットの出力層にソフトマックス関数という離散的なクラス分類に用いる関数を使用しているが、これを連続的な出力にした場合予測と正答の誤差が縮まるのではないかと考えられる。

しかし課題としていかなる出力層・活性化関数が適当なのかという問題と、出力が連続値の場合、単純に正解・不正解を区別することが出来ないため、いかなる方法によってモデルの精度を評価したり、他の機械学習手法と比較したりするかという問題が有り、定性的・定量的な評価基準を新たに開発しなければならない。

第B.5.3項　　RNN 及びその派生手法にかかる時間の問題

第 B.3 節において GRU を用いたアンケートデータの推定を行ったが、同節においてテストデータの抜き出し回数は *XX* 回であり、第 B.2 節の実験におけるテストデータ抜き出し 50 回と平均正解率と定性的な比較をすることが出来ない。しかし GRU の推定では第 5.2 節の圧縮作業を行う前のデータ約 30 万個を使用するため、テストデータ抜き出し 1 回あたりの計算時間が 7 ～４８時間と極めて長く抜き出しを５０回行うことが時間的な問題で困難であると考えられたためである。また、計算時間は GRU のユニット数・層数ともやや比例しているため、複雑な分析を行わせようとすると極めて時間がかかる。

このため、GRU 投入前に何らかの方法で入力データを削減・圧縮するか計算時間を短縮する方法を考える必要がある。

第B.5.4項　NN 出力層におけるクラス数問題

第 5.4 節～第 5.5 節及び第 B.2 節～第 B.3 節の実験では、いずれも NN の出力層にソフトマックス関数を用い 5 クラス分類問題として解いている。第 B.5.2 項でも述べたとおり、誤差範囲が 1 である「惜しい」データが多い事から、クラス分けを減らすことで相対的に正解率を上げることができるのではないかと考えられる。当初の研究目的通り、生徒の学習作業中の心理変化のタイミングを検出することが目的ならば、心理変化の程度を厳密に測定する必要はなく、2 ～ 3 クラス程度で十分と思われる。