

深層学習を利用した脳波情報に基づく音楽推薦システム

鷲沢研究室

1631014

糸賀 弘樹

指導教員：鷲沢 嘉一 准教授

2018 年 1 月 29 日

目次

第 1 章	序論	4
1.1	はじめに	4
1.2	本論文で用いる記号表記	5
第 2 章	従来手法	6
2.1	脳波	6
2.1.1	周波数解析	6
2.2	ランキング学習	7
2.2.1	ポイントワイズランキング学習	7
2.2.2	ペアワイズランキング学習	7
2.3	サポートベクトルマシン (SVM)	8
2.3.1	サポートベクトルマシン (SVM)	8
2.3.2	ランキングサポートベクトルマシン (Ranking SVM)	11
2.3.3	サポートベクトル回帰	11
2.4	ニューラルネットワーク	12
2.4.1	深層学習	12
第 3 章	提案手法	16
3.1	ランキングディープニューラルネットワーク	16
3.2	脳波からの音楽推薦システム	16
第 4 章	実験	19
4.1	実験	19
4.1.1	実験環境	19
4.1.2	実験方法	19
4.1.3	解析手法	19
4.1.4	データ整形とセッティング	20
4.2	結果	21
4.2.1	normalized Discounted Cumulative Gain (nDCG)	21
4.2.2	順位付け精度	22
4.2.3	DRNN のパラメータの検討	22

4.2.4	特徴量の検討	23
4.3	考察	23
第 5 章	結論	32
5.1	まとめと今後の課題	32
参考文献		34

第 1 章

序論

1.1 はじめに

インターネット関連の技術の発達により、現代社会には無数のデジタルコンテンツ (アイテム) が存在する。これら全てのアイテムを、我々が利用することは難しく、多くの場合はその中から自分自身の目的や好みに適したアイテムを探し出し、利用する。しかしながら、無数のアイテムから自力で探し出すことは、多くの時間やコストを必要とする。そのため、我々は、他者からの推薦に頼ることで、必要とされる時間やコストを少なくし、目的のアイテムを探す。推薦を行う他者は、必ずしも人ではなく、機械が行う場合もある。例えば、Amazon のウェブサイトでは、利用者の購買履歴や、チェックした商品などの情報を元に、アイテムの推薦を行う。このようなシステムは、機械学習を用いており、推薦システムと呼ばれている。

機械学習とは、与えられる訓練データから、ルールや基準を獲得していく手法である。機械学習が扱う問題は大きく教師あり学習、教師なし学習と、強化学習の 3 つに分類される。教師あり学習は、訓練データに対してラベル付けが行われたデータを扱い、ラベルの情報と、訓練データの特徴量から学習を行う。一方、教師なし学習は、ラベル付けが行われていない訓練データを用いて学習を行い、背後の規則を獲得する。機械学習の代表的な手法として、深層学習がある [1]。深層学習は、自己符号化器を多層にすることで、従来のニューラルネットワークが抱えていた勾配消失問題を解決し、様々な分野で高い性能を示してきた。

現在用いられている推薦システムの多くは、利用者により明示的に与えられるアイテムに対するレーティングや、非明示的に与えられる購買履歴などから作成されたモデルに基づいて推薦を行う。しかし、レーティングや購買履歴により作成されたモデルは、虚偽のモデルが簡単に作成可能である問題がある。この問題点を用いた攻撃はサクラ攻撃と呼ばれる。サクラ攻撃は、雇った人間や、プログラムにより、自身に有利な推薦が行われるよう推薦システムへ情報を入力させる行為である。サクラ攻撃に対して、様々な防御方法が提案されているが、ユーザが虚偽の情報を入力することができないシステムを構築することにより、この問題は解決可能であると考えられる。

信号の計測及び、処理技術の発展に伴い、ここ数十年の間に生体信号を扱った研究が盛んに行われている。センサや処理デバイスの小型化、高性能化に伴い、生体信号の現実社会での利用も徐々に行われている。生体信号は心拍や脳波などの生体現象により生じる信号である。人間は、意識的に内面的変化を隠すことができるが、生体信号を偽ることは困難であることから、生体信号が様々な評

価指標として用いられている。例えば、生体信号を用いて、個人認証を行う研究がある。この研究は、人間の生物学的個性を自動識別し、個人認証を行う。複数の生体信号を組み合わせることにより、90%以上の識別精度が示された [2]。生体信号の中でも、脳は精神や意識などを支配しているため、それらの変化は、脳波に現れるという考えから、心理状態の変化と脳波に注目する研究が多くある [3]。脳波から人間の感情状態を推定する研究では、音楽鑑賞時の脳波から感情の4クラス分類を行い、82.29%の識別率を示した [4]。脳波から感性や好みを抽出する研究では、好みの匂いと嫌いな匂いを嗅いだ際の脳波から、好みについての情報を抽出する手法が提案された [5]。これらの研究から、脳波から得られる情報の信頼性が保障されたため、脳波からモデルを作成し、そのモデルに基づき行われる推薦は、従来のシステムより、強固で正確であることが期待できる。

好みの推定は、定量的に好みを推定する手法と、好みの順位付けを推定する手法がある。定量的な好みの推定は、回帰を用いることで実現可能である。これまでの研究では、サポートベクター回帰を用いて脳波から楽曲の好みの10段階評価を推定したが、回帰誤差が大きく実用が困難なことが示された [6]。一方、順位付けを推定する手法はランキング学習と呼ばれ、与えられた二つ以上のデータのうちどちらがより上位であるかを学習する。ランキング学習は、単一のデータに対して評価を行うポイントワイズと、二つのデータの比較を行うペアワイズがある。定量評価と順位付けの二つの手法の違いは、定量推定では好みの度合いが重要であるのに対して、順位付けでは、どちらがより好みであるかのみに注目されることである。

本研究の目的は、脳波から音楽の好みを推定し、得られた好みの情報に基づいた推薦システムを実現することである。脳波から得られる情報は、偽ることができないため、従来のシステムと比較し、信頼ができるシステムとなることが期待出来る。このため、好みの推定に用いたランキング学習の手法について、深層学習を用いた手法を提案する。サポートベクター回帰 (SVR)、ランキングサポートベクターマシン (RSVM) [7] や深層学習 (ANN) によるポイントワイズ、ペアワイズの順位付けの手法の比較検討を行う。提案したシステムの有効性を確かめるため、脳波測定の実験を行い、データの解析を行った。

1.2 本論文で用いる記号表記

本論文では、数字とスカラー定数はローマン体、スカラー変数はイタリック体、ベクトルと行列はボールドイタリック体で表す。

第 2 章

従来手法

2.1 脳波

センサ技術や、信号処理技術の発展に伴い、生体信号の研究が盛んに行われている。特に、脳が精神や意識を支配しているため、精神や意識の変化は脳波に現れるという考えに基づいて、様々な研究が行われてきた。脳波とは、脳内の神経細胞の活動により生じる微小な電気活動を、頭表上の電極から測定したものである [8]。脳波は電位の発生源を取り囲む電導性生体組織（脳、脳髄液、血管、頭骨、頭皮）の外側から測定したもので、侵襲性は低いが多数の細胞の活動集合が記録されるという限界がある。このため、脳波から取り出したい情報があるとき、脳波に対して適切な処理を行い、含まれた関心のある情報のみを取り出す必要がある。また、頭表から測定される電位は数 μV 程度である。この微小な電位の測定には差動増幅を行う。差動増幅は基準電極の電圧と計測点の電圧の差をオペアンプにより増幅し、微小な電圧変化を計測する手法である。脳波を測定する際には一般的に耳たぶや眉間の間に基準電極を配置する。電極は一般的に、国際 10-20 システム (図 2.1) または、それを拡張した拡張国際 10-20 システムに従って配置する。

2.1.1 周波数解析

脳波が持つ情報として周波数は重要であり、周波数解析はよく行われる解析方法である。脳波の周波数成分で最も顕著なものは 8~13Hz の成分で α 波と呼ばれる [10]。この α 波は静かな暗室内で目を閉じなるべく何も考えていないようにしていると最も規則的に大きな振幅で現れる。しかし、物を見つめたり考えたりすることにより α 波が減少し、不規則な細かい波が目立つようになる。これは α 波減衰とよばれる現象で、この状態のときに相対的に目立つ 13Hz 以上の成分を β 波という。また、成人男性において、特に睡眠時には 0.5~8Hz の低周波成分が目立つようになる。これらは δ 波と θ 波と区別される。 α 波や θ 波は感情状態との関連が報告されてる。[11, 12] このように脳波は感覚入力、運動、覚醒状態の変化、認知活動などによって周波数成分が変わることが知られている。各周波数帯域の範囲については大まかな基準しか決まっておらず、本研究では 4~8Hz を θ 波、8~10Hz を low- α 波、10~13Hz を high- α 波、13~20Hz を low- β 波、20~30Hz を high- β 波と分類する。

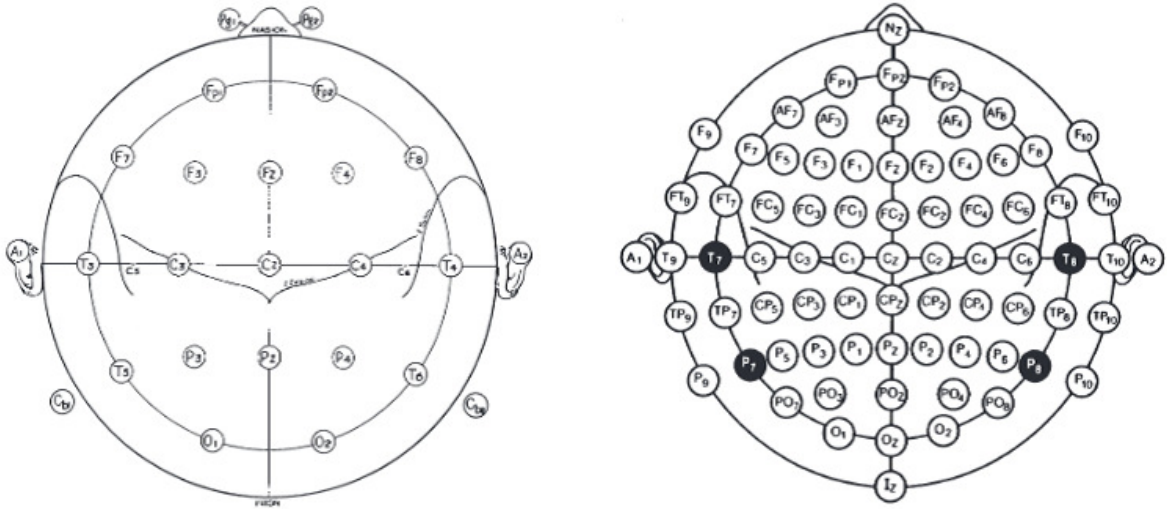


図 2.1 国際 10-20 システム (左) と拡張国際 10-20 システム (右) (出展 [9])

2.2 ランキング学習

ランキング学習とは、アイテムの順位付けの方法を獲得する手法である。web 技術の発展により、増加した文章を人が利用者に対して提案することは困難になり、機械による推薦が求められた。そのため、機械学習の手法を用いた様々なランキング学習の手法が提案されてきた [13, 14, 15]。代表的な手法は、ポイントワイズとペアワイズに分けることができる。ポイントワイズは 1 つのアイテムに対する評価から順位付けを行い、ペアワイズは 2 つのアイテムの差分から順位付けを行う。

2.2.1 ポイントワイズランキング学習

ポイントワイズでは、1 つのアイテムに対する評価から順位付けを行う。アイテムから得られた特徴 \mathbf{x} と、目的変数 y が与えられたとき、重み \mathbf{w} とマッピング関数 $\phi(\cdot)$ を用いて、

$$y = \mathbf{w}^T \phi(\mathbf{x}) \quad (2.1)$$

となる \mathbf{w} を学習する。テストデータに対して、同様の重みを用いて y を推定し、その大きさによって順位付けを行う。

2.2.2 ペアワイズランキング学習

ペアワイズ手法では、2 つのアイテムのペアを扱う。2 つのアイテムから得られた特徴 $(\mathbf{x}_a, \mathbf{x}_b)$ と、それぞれの目的変数 (y_a, y_b) が与えられたとする。ただし、 $(y_a \neq y_b)$ である。新たな特徴ベクトル \mathbf{x} を、

$$\mathbf{x} = \mathbf{x}_a - \mathbf{x}_b \quad (2.2)$$

とし、そのラベル y を

$$y = H(y_a - y_b) \quad (2.3)$$

とする。ここで $H(\cdot)$ はヘヴィサイドのステップ関数である。重み \mathbf{w} とマッピング関数 $\phi(\cdot)$ を用いて、

$$y = \begin{cases} 1 & (\phi(\mathbf{w}^T \mathbf{x}) > 0) \\ 0 & (\text{otherwise}) \end{cases} \quad (2.4)$$

となる \mathbf{w} を 2 値分類問題と同様の手法で学習する。テストデータに対しては、同様の重みを用いて得られた y が 2 つのデータの順位付けを表しているため、その情報をもとに全体の順位付けを行う。

2.3 サポートベクトルマシン (SVM)

2.3.1 サポートベクトルマシン (SVM)

サポートベクトルマシン (SVM; Support Vector Machine) は、Vapnik らにより開発された 2 クラス分類問題の代表的手法であり、未知データに対して高い予測精度を持つことが知られている [16, 17].

2.3.1.1 2 クラス分類問題

2 クラス分類問題とは、与えられた入力データが 2 つのクラスのどちらに属するかを分類する問題を考える問題である。クラスを推定する処理を行うものを識別器と呼ぶ。識別器は事前に訓練データと呼ばれる分類対象の学習データが与えられる。よって分類問題は訓練データから分類のための規則を獲得する問題といえる。 n 個の訓練データを $\{(\mathbf{x}_i, y_i)\}_{i \in [n]}$ と表す。 $i \in [n]$ は $i = 1, 2, 3, \dots, n$ を表す。ここで \mathbf{x}_i を特徴ベクトル、 y_i をラベル、ある \mathbf{x}_i と y_i の組を事例と呼ぶ。識別器はある \mathbf{x} が与えられたとき、その y の予想値を返す関数であると解釈できる。

2.3.1.2 ハードマージン SVM

n 個の事例からなる訓練集合 $\{(\mathbf{x}_i, y_i)\}_{i \in [n]}$ が d 次元実数ベクトル $\mathbf{x}_i \in \mathbb{R}^d$ と 1 か -1 の値をとるラベル $y_i \in \{-1, 1\}$ から構成されているとする。決定関数と呼ばれる実数値関数 $f: \mathbb{R}^d \rightarrow \mathbb{R}$ を用いて次のように識別器 $g(\mathbf{x})$ を定義する。

$$g(\mathbf{x}) = \begin{cases} 1 & (f(\mathbf{x}) > 0) \\ -1 & (f(\mathbf{x}) < 0) \end{cases} \quad (2.5)$$

ここでは $f(\mathbf{x})$ として以下の一次関数を考える。

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (2.6)$$

ここで、 $\mathbf{w} \in \mathbb{R}^d$, $b \in \mathbb{R}$ である。訓練集合内の全てのデータが線形分離可能であるならば全ての i に対して $y_i f(\mathbf{x}_i) > 0$ となる $f(\mathbf{x})$ が存在する。一般にこの $f(\mathbf{x})$ は複数存在するが、SVM においてはそれぞれのクラスのデータが識別の境界である $f(\mathbf{x}) = 0$ からなるべく離れるように定める。データ

が境界からどのくらい離れているかをマージンと呼ぶ, SVM はマージンを最大化する $f(\mathbf{x})$ を求める. \mathbf{x}_i から分類境界までの距離は次のように表される.

$$\frac{|\mathbf{w}^T \mathbf{x}_i + b|}{\|\mathbf{w}\|} \quad (2.7)$$

全ての事例を正しく識別し, マージンを最大化する最適化問題は次のように表せる.

$$\begin{aligned} \max_{\mathbf{w}, b} \quad & \frac{1}{2} \frac{M}{\|\mathbf{w}\|^2} \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq M, i \in [n] \end{aligned} \quad (2.8)$$

制約条件より, M は全ての事例に対する $y_i(\mathbf{w}^T \mathbf{x}_i + b)$ の値のうち最も小さい値である. $\mathbf{w} \leftarrow \mathbf{w}/M$, $b \leftarrow b/M$ と定義しなおし, (2.8) は次のように表せる.

$$\begin{aligned} \max_{\mathbf{w}, b} \quad & \frac{1}{2} \frac{1}{\|\mathbf{w}\|} \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i \in [n] \end{aligned} \quad (2.9)$$

$1/\|\mathbf{w}\|$ の最大化は $\|\mathbf{w}\|$ の最小化と等価であるため, マージン最大化は次の最適化問題で表せる.

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i \in [n] \end{aligned} \quad (2.10)$$

このような SVM はハードマージン SVM と呼ばれる. (2.10) の最適解を求めると, $y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$ を満たす \mathbf{x}_i が一般にいくつか現れる. これは分類境界に最も近い \mathbf{x}_i であり, サポートベクトルと呼ばれる.

2.3.1.3 ソフトマージン SVM

ハードマージン SVM では訓練事例全てが線形分離可能である必要があるが, 実際の問題ではそうでないことが多い. よって制約条件を緩和することにより, 線形分離不可能なデータに対応する. 新たに非負の実数 $\xi_i \geq 0, i \in [n]$ を導入し, (2.10) を以下のように変形する.

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \in [n]} \xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i \in [n] \end{aligned} \quad (2.11)$$

(2.11) の定数 C はコストパラメータと呼ばれ, 事前に設定が必要なハイパーパラメータである. C が大きいほどハードマージン SVM に近づき, $C \rightarrow \infty$ のときハードマージン SVM と一致する.

2.3.1.4 双対問題とカーネル

(2.10) や (2.11) といった最適化問題は主問題と呼ばれる。これに対し，双対問題と呼ばれる問題を解くことで，同じ最適化問題に対して違った見方が得られる。(2.11) を次のように書き換える。

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \in [n]} \xi_i \quad (2.12)$$

$$s.t. \quad -1(y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i) \leq 0, -\xi_i \leq 0, i \in [n]$$

ここで新たに $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)^T$ と $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ を導入し，(2.13) を次のような最適化問題に変形する。

$$\min_{\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\mu}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i \in [n]} \xi_i - \sum_{i \in [n]} \alpha_i (y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i) - \sum_{i \in [n]} \mu_i \xi_i \quad (2.13)$$

(2.13) の解は Karush-Kuhn-Tucker (KKT) 条件を満足することが知られているため，(2.11) の双対問題は次のように示される。

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & -\frac{1}{2} \sum_{i, j \in [n]} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i \in [n]} \alpha_i \\ s.t. \quad & \sum_{i \in [n]} \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i \in [n] \end{aligned} \quad (2.14)$$

双対問題を解き，得られた $\boldsymbol{\alpha}$ から $f(\mathbf{x})$ は KKT 条件のひとつである $\mathbf{w} = \sum_{i \in [n]} \alpha_i y_i \mathbf{x}_i$ より，次のように表現できる。

$$f(\mathbf{x}) = \sum_{i \in [n]} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \quad (2.15)$$

b は $\boldsymbol{\alpha}$ を用いて次のように計算される。

$$b = y_i - \sum_{i' \in [n]} \alpha_{i'} y_{i'} \mathbf{x}_{i'}^T \mathbf{x}_i, i \in \{j \in [n] | 0 < \alpha_j < C\} \quad (2.16)$$

ここで入力 \mathbf{x} を何らかの特徴空間 \mathcal{F} へ写像する関数 $\phi: \mathbb{R}^d \rightarrow \mathcal{F}$ を考える。ここで $\phi(\mathbf{x})$ を新たな特徴ベクトルと解釈し， $f(\mathbf{x})$ を次のように変化させる。

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (2.17)$$

ϕ が非線形であれば $f(\mathbf{x}) = 0$ によって定義される識別の境界はもとの \mathbf{x} の空間では非線形となる。一方，空間 \mathcal{F} では $f(\mathbf{x})$ は線形な分離境界を持つ。よって $\phi(\mathbf{x})$ を新たな特徴ベクトルとみなすことで次式が得られる。

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & -\frac{1}{2} \sum_{i, j \in [n]} \alpha_i \alpha_j y_i y_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) - \sum_{i \in [n]} \alpha_i \\ s.t. \quad & \sum_{i \in [n]} \alpha_i y_i = 0 \end{aligned} \quad (2.18)$$

$$0 \leq \alpha_i \leq C, i \in [n]$$

ϕ は内積の形で現れているため, $\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ をカーネル関数として次のように定義する.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (2.19)$$

これを用いて式 (2.18) は次式で表される.

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i,j \in [n]} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i \in [n]} \alpha_i \\ \text{s.t.} \quad & \sum_{i \in [n]} \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i \in [n] \end{aligned} \quad (2.20)$$

また, $f(\mathbf{x})$ は次式で表される.

$$f(\mathbf{x}) = \sum_{i \in [n]} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (2.21)$$

これにより非線形分類を行う SVM が実現できる.

2.3.2 ランキングサポートベクトルマシン (Ranking SVM)

ランキングサポートベクトルマシンは, ペアワイズのランキング学習手法である [7]. N 組の訓練データ $(\mathbf{x}_i, y_i)_{i=1}^N$ が与えられたとき, 不等式

$$y_i > y_j \Rightarrow \mathbf{w}^T \phi(\mathbf{x}_i) > \mathbf{w}^T \phi(\mathbf{x}_j) \quad (2.22)$$

を満たす重み \mathbf{w} を求める. Ranking SVM では, 次の最適化問題の解として \mathbf{w} が与えられる.

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i,j} \xi_{ij} \\ \text{s.t.} \quad & \mathbf{w}^T (\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)) \geq 1 - \xi_{ij}, \quad \xi_{ij} \geq 0, \quad \forall i \neq j \end{aligned}$$

ここで, ϕ はマッピング関数, ξ はスラック変数, C ($0 < C < \infty$) はコストパラメータである. 最適化問題は SVM と同様の手法により解くことが可能である.

2.3.3 サポートベクトル回帰

サポートベクトル回帰は, SVM の回帰問題への拡張である [18]. 回帰問題では, N 個の特徴ベクトル $\mathbf{x}_i (i = 1, 2, \dots, N)$ が与えられたとき, 出力 y_i は次のように求まる.

$$y_i = \mathbf{w}^T \phi(\mathbf{x}_i) + b \quad (2.23)$$

ここで, $\phi()$ はマッピング関数, \mathbf{w} は重み, b はバイアスである. マッピング関数が線形であれば線形回帰となり, 非線形であれば非線形回帰となる. このとき, 特徴ベクトル \mathbf{x} に対応したラベル t と出

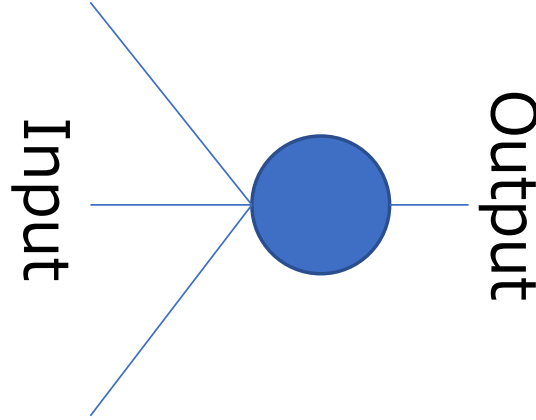


図 2.2 ニューロン

力の誤差が小さくなるようなパラメータを求める．単純に誤差を最小にする場合に，過適合などの問題が生じる可能性があるため，サポートベクトル回帰では $\epsilon > 0$ に対して次のように定式化される．

$$\begin{array}{ll} \min & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i (\xi_i + \xi_i^*) \\ \text{s.t.} & \mathbf{w}^T \phi(\mathbf{x}_i) + b - y_i \leq \epsilon - \xi_i \\ & y_i - \mathbf{w}^T \phi(\mathbf{x}_i) - b \leq \epsilon - \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \dots, N \end{array}$$

ここで， C は正則化パラメータ， ξ_i, ξ_i^* はスラック変数である．最適化問題は SVM と同様の手法により解くことが可能である．

2.4 ニューラルネットワーク

ニューラルネットワークは，非線形のモデルの一つで，人間の神経細胞を模したニューロン (図 2.2) を複数個連結したものである．一般的なニューラルネットワークは図 2.3 のように 3 つの層からなり，それぞれ入力層，中間層，出力層と呼ばれる． n 個の入力層のユニットと m 個の出力層のユニットを持つ順伝播型ニューラルネットワークを考える．このとき，出力層からの出力 \mathbf{y} は入力 \mathbf{x} を用いて次式のように表される．

$$\mathbf{y} = f_2(\mathbf{W}_2 f_1(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) \quad (2.24)$$

ここで， $f_1, f_2(\cdot)$ は活性化関数， $\mathbf{W}_{1,2}$ は重み， $\mathbf{b}_{1,2}$ はバイアスと呼ばれる．活性化関数は非線形関数，特に Rectified Linear Unit (ReLU) やシグモイド関数が用いられる [19]．各ニューロンはユニットと呼ばれ，図 2.4 に示されるように，ユニットの出力は，次の層のユニットへそれぞれ異なる重みを掛けられ，伝播する．ユニットに入力された前層のユニットたちの出力は，バイアスと足し合わされ，活性化関数により，非線形変換が行われる．

2.4.1 深層学習

深層学習とは，中間層の数が 2 以上のニューラルネットワークを用いて学習を行うことである．従来の多層のニューラルネットワークは，パラメータの更新を行う際に用いられる勾配が，シグモイド

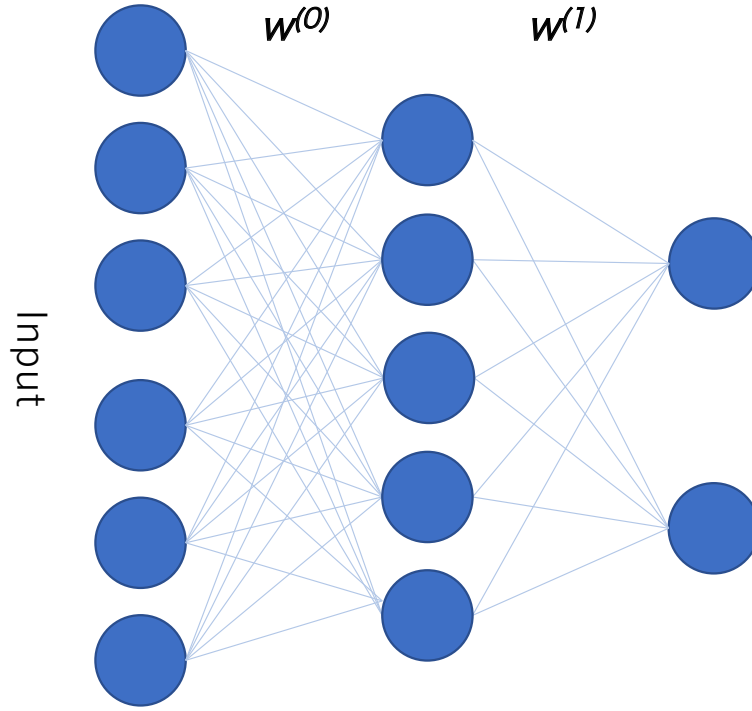


図 2.3 ニューラルネットワーク

関数などの活性化関数を多数介することにより消失してしまう問題を抱えていた．これを解決するため，Hinton らは，自己符号化器を提案した [1]．深層学習は，多くの分野において，従来の機械学習手法を超える結果を示した [20, 21]． L 個の層を持つ，多層のネットワークについて考える． $l, l+1$ 層のユニット数をそれぞれ， I, J とする． l 層から $l+1$ 層への伝播は， l 層の出力を $\mathbf{z}^{(l)}$ として，

$$\mathbf{z}^{(l+1)} = f(\mathbf{W}^{(l+1)} \mathbf{z}^{(l)} + \mathbf{b}^{(l+1)}) \quad (2.25)$$

と書ける．ニューラルネットワークにおけるパラメータの更新の指標として，誤差関数が用いられる．重みとバイアスをまとめて \mathbf{w} とする．誤差関数 $E(\mathbf{w})$ に対し，最小値を与える $\mathbf{w} = \arg \min_{\mathbf{w}} E(\mathbf{w})$ を求める．しかし，一般に $E(\mathbf{w})$ は凸関数でないため，最小の解を求めることは難しい．よって，代わりに $E(\mathbf{w})$ の極小点を求める．極小点を反復計算により求める方法として勾配降下法がある．勾配とは以下の式で表されるベクトルである．

$$\nabla E = \frac{\partial E}{\partial \mathbf{w}} = \left[\frac{\partial E}{\partial \mathbf{W}} \quad \frac{\partial E}{\partial \mathbf{b}} \right]^T \quad (2.26)$$

勾配降下法は，勾配を用いて以下の式に従いパラメータの更新を行う．

$$\mathbf{w}^{\text{new}} = \mathbf{w}^{\text{old}} - \epsilon \nabla E \quad (2.27)$$

ここで， ϵ は，学習率と呼ばれ，更新量の大きさを決定するパラメータである．

局所的な極小値への収束を避ける方法として，確率的勾配降下法がある．勾配降下法では，誤差は全てのサンプルの誤差関数の和として計算されるが，1 つのサンプル $E_n(\mathbf{w})$ の誤差を用いてパラメータを更新する．

$$\mathbf{w}^{\text{new}} = \mathbf{w}^{\text{old}} - \epsilon \nabla E_n \quad (2.28)$$

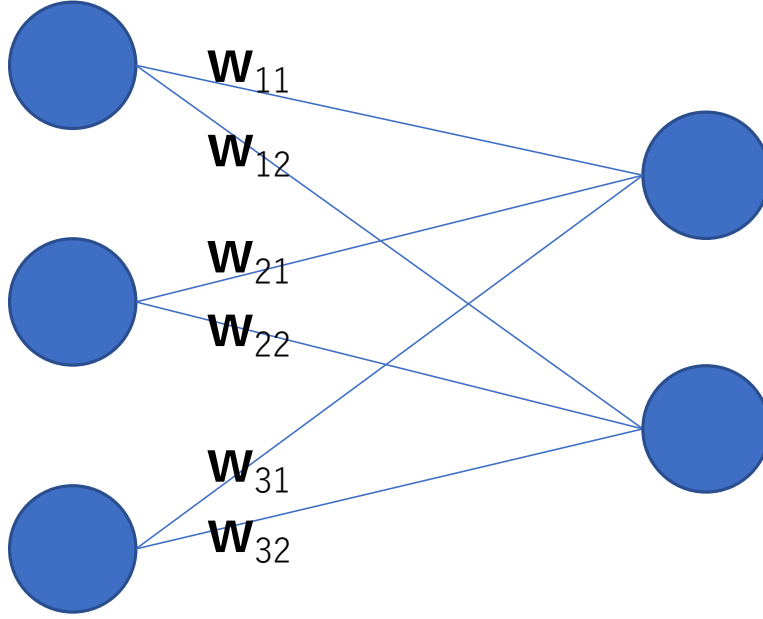


図 2.4 順伝播

このとき、サンプルは更新のたびにランダムに選択される。また、勾配降下法と確率的勾配降下法の間に当たる、ミニバッチと呼ばれる手法がある。これは、勾配計算を少数のデータを用いて行う手法である。

深層学習においては、誤差関数の計算に多くのコストがかかる。計算コスト削減のために、誤差逆伝播法が広く用いられている。 $l-1$ 層目の i 番目のユニットから l 層目のユニットへの伝播にかかる重み $w_{ji}^{(l)}$ に関する 1 つのサンプルの誤差関数の勾配 $\frac{\partial E_n}{\partial w_{ji}^{(l)}}$ について考える。 $w_{ji}^{(l)}$ は次の $u_j^{(l)}$ の順伝播にのみ用いられる。

$$u_j^{(l)} = \sum_i w_{ji}^{(l)} z_i^{(l-1)} \quad (2.29)$$

ここで $z_i^{(l-1)}$ は $l-1$ 層目の i 番目のユニットの出力である。連鎖規則により、

$$\frac{\partial E_n}{\partial w_{ji}^{(l)}} = \frac{\partial E_n}{\partial u_j^{(l)}} \frac{\partial u_j^{(l)}}{\partial w_{ji}^{(l)}} \quad (2.30)$$

が成り立つ。 $u_j^{(l)}$ は $l+1$ 層の全てのノードに重みがかけられ伝播するため、

$$\frac{\partial E_n}{\partial u_j^{(l+1)}} = \sum_k \frac{\partial E_n}{\partial u_k^{(l+1)}} \frac{\partial u_k^{(l+1)}}{\partial u_j^{(l)}} \quad (2.31)$$

と書ける。ここで、 l 層 j 番目に関する誤差の微分を

$$\delta_j^{(l)} \equiv \frac{\partial E_n}{\partial u_j^{(l+1)}} \quad (2.32)$$

とすると,

$$\delta_j^{(l)} = \sum_k \delta_k^{(l+1)} (w_{kj}^{(l+1)} f(u_j^{(l)})) \quad (2.33)$$

と書けることから, $l+1$ 層の誤差の微分により, l 層の誤差の微分が計算できることがわかる. ここで,

$$\frac{\partial u_j^{(l)}}{\partial w_{ji}^{(l)}} = z_i^{(l-1)} \quad (2.34)$$

であるから, 誤差関数の勾配は

$$\frac{\partial E_n}{\partial w_{ji}^{(l)}} = \delta_j^{(l)} z_i^{(l-1)} \quad (2.35)$$

のように表される.

第3章

提案手法

3.1 ランキングディープニューラルネットワーク

深層学習は様々な分野で高い性能を発揮していることから、ランキング学習においても高い性能が期待できる。このため、深層学習をペアワイズのランキング学習と組み合わせたランキングディープニューラルネットワークを3つ提案する。

1つ目のネットワーク (RDNN1) を図 3.1 に示す。Data A と Data B はペアワイズ学習を行う2つの入力データである。RDNN1 は第1層目でランキングサポートベクトルマシンのように各データからの特徴抽出を行う。重み \mathbf{W}_1 は2つの入力データで共通した値をとるため、入力の順序を入れ替えたとしても同様の特征抽出が行われる。第2層目では、重みを ± 1 または 0 とすることで2つの入力データの対応した特徴量の差をとる。第2層目の重みは学習時に更新されることはない。得られた差を一般の深層学習の入力として扱い、順伝播させ出力を得る。

2つ目のネットワーク (RDNN2) を図 3.2 に示す。RDNN2 は RDNN1 に課した第2層における重みに関する制約を緩めたネットワークである。重み \mathbf{W}_2 の初期値を、対応する特徴同士は ± 1 、そうでない重みは絶対値が 1 より小さい値とし、学習により重みが更新されるようにした。

3つ目のネットワーク (RDNN3) を図 3.3 に示す。RDNN3 は重みに関する全ての制約をなくしたネットワークであり、一般的な深層学習の入力に2つの入力データを入れている。制約を無くすることで、ネットワークの表現能力は向上するが、誤差関数が大域的な最小値でない局所解に収束する可能性が高くなる。これらのネットワークは、ランキングサポートベクトルマシンと同様にどちらの入力が高いランクであるかを示す。

本研究において、ネットワークの層数は6層とした。第2層と出力層以外の活性化関数は、Rectified Linear Unit (Relu)、第2層では恒等関数、出力層ではロジスティックシグモイド関数を用いた。固定されていない重みは、自己符号化器を用いて初期化を行い、誤差逆伝播法によりファインチューニングを行った。パラメータの更新は確率的勾配降下法、または、ミニバッチを用いた。

3.2 脳波からの音楽推薦システム

提案する音楽推薦システムの概要を図 3.4 に示す。

音楽聴取時の脳波の測定を行い、特徴抽出を行う。ランク付けのための学習は、ラベル付けされた

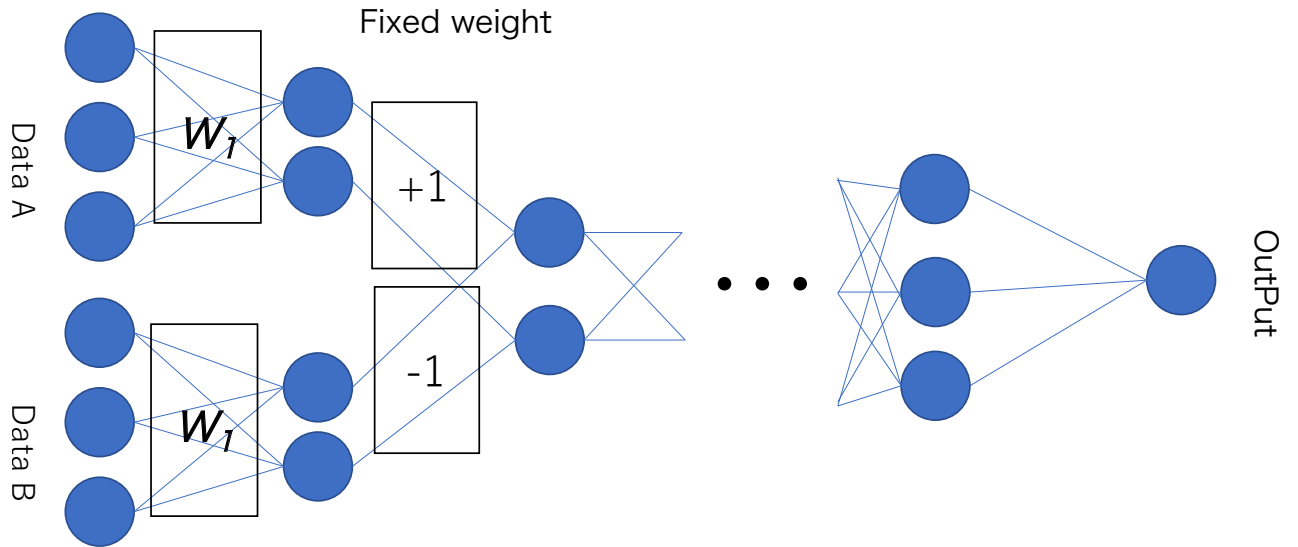


図 3.1 ランキングディープニューラルネットワーク 1

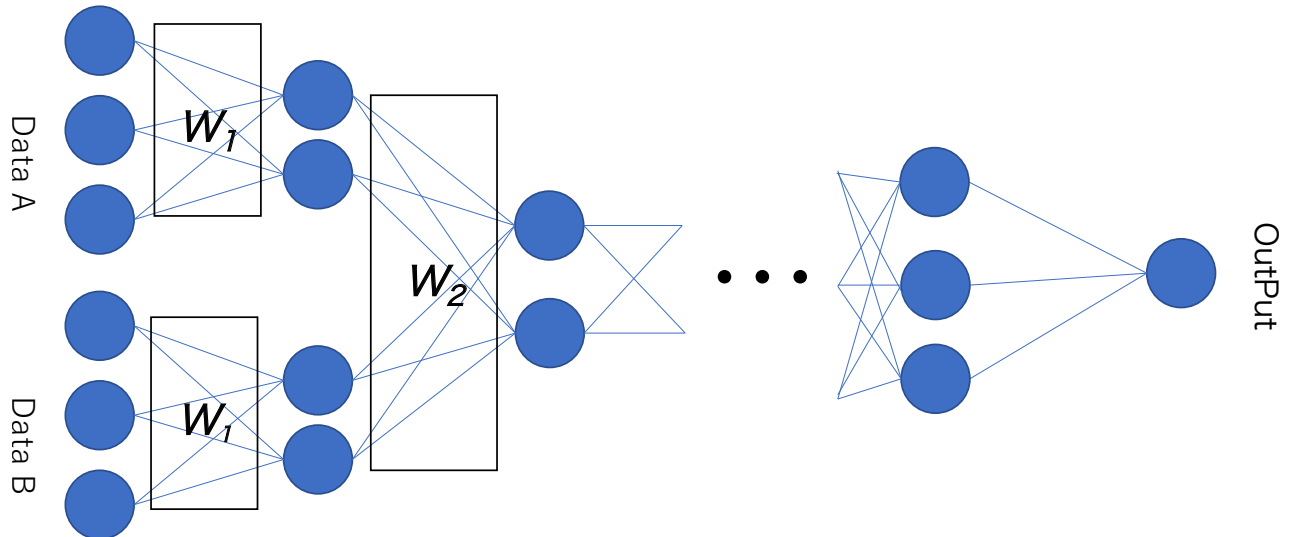


図 3.2 ランキングディープニューラルネットワーク 2

特徴から行い，得られた重みとテストデータから，利用者に対して視聴した音楽のランク付けの結果を提示する．被験者がつけたスコアを推定するのではなく順位をペアワイズの手法で学習し推定することで，データ収集が困難な脳波データ数を増やすことが可能である．また，実際にテストを行う段階で順位付けを評価することは，スコアの評価よりも正確であることが期待出来る．システムの問題点として，音楽鑑賞時の脳波が存在しない場合推薦を行うことができないことがあげられる．しかし，この問題は協調フィルタリングを用いることで解決できると考えられる．

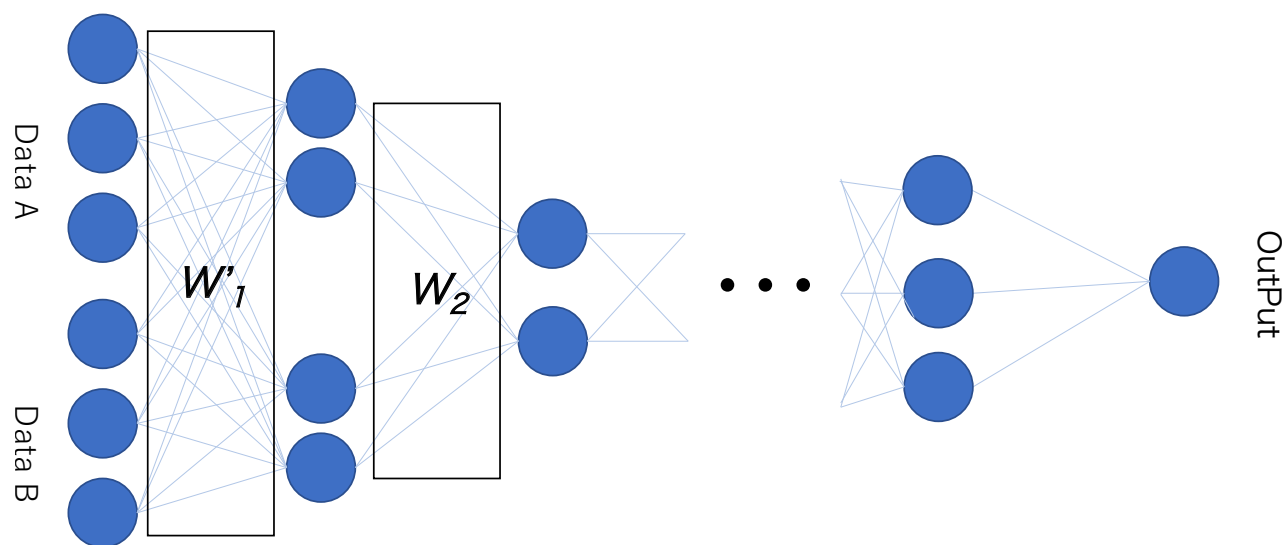


図 3.3 ランキングディープニューラルネットワーク 3

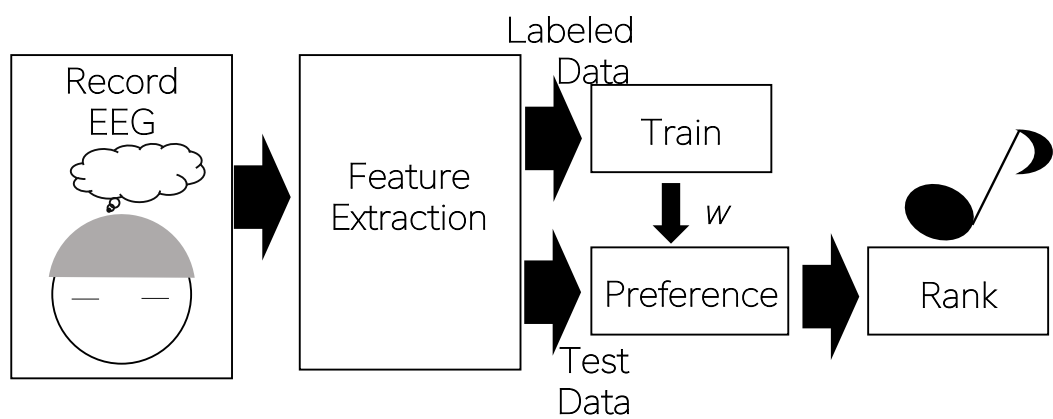


図 3.4 Music Recommendation System

第 4 章

実験

4.1 実験

提案したランキング学習手法の有効性を確かめるために実験を行った。

4.1.1 実験環境

実験は 5 人の健康な 20 代男性に対して行った。脳波は Guger technologies 製のアクティブ EEG を用いて測定し、電極を拡張国際 10-20 法に従って Fp1, Fp2, F5, Fz, F6, FT7, FT8, C5, Cz, C6 に配置した。リファレンスは Fpz, グランドは A2 に配置した。電極配置図を図 4.1 に示す。サンプリング周波数は 512Hz とし、測定後に 128Hz ヘダウンサンプリングした。

4.1.2 実験方法

全ての被験者は脳波測定中、椅子に座り目を瞑りリラックスした状態を維持した。初めに被験者は 15 秒間の切り出された音楽を聴き、その後聴いた音楽を 1 (嫌い) から 10 (好き) の 10 段階で評価した。音楽を聴き、評価を行うことを 100 回繰り返す。全ての被験者から 100 のデータを得た。その中から、 $60\mu V$ 以上の電圧が記録されていたデータをアーチファクトノイズが記録されていたデータとし、除去を行った。

4.1.3 解析手法

脳波を解析するため、パワースペクトル密度, differential asymmetry (DASM) を特徴量として抽出した。音楽鑑賞時の心理状態の時間変化を捉えるため、2 秒毎と曲全体を通したパワースペクトル密度の 2 つを用いた。測定した脳波を $x[n]$ とし、離散フーリエ変換を適用する。

$$X[k] = \sum_{n=0}^{N-1} x[n] \exp(-j \frac{2\pi kn}{N}) \quad (4.1)$$

ここで、 N は標本点の数である。これを用いて PSD は次のように表される。

$$P[k] = |X[k]|^2 \quad (4.2)$$

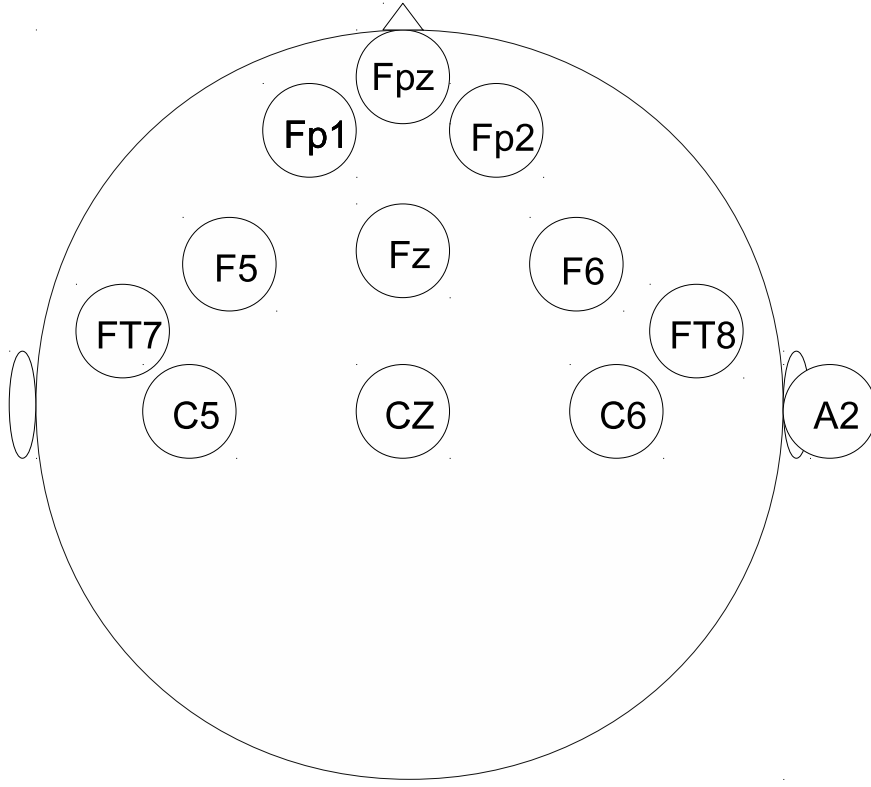


図 4.1 電極配置図

前に述べた θ 帯域, low- α 帯域, high- α 帯域, low- β 帯域, high- β 帯域の 5 つの帯域におけるパワーが特徴として用いられる. 特定の周波数帯域, $k_1, k_2 (0 \leq k_1 \leq k_2 \leq n)$ におけるパワー $P_{k_1 k_2}$ は $P(k)$ を用いて, 次のように表される.

$$P_{k_1, k_2} = \sum_{k=k_1}^{k_2} P[k] \quad (4.3)$$

例えば, θ 帯域のパワーは,

$$P_\theta = P_{4,8} = \sum_{k=4}^8 P[k] \quad (4.4)$$

と表される. DASM は脳波からの感情推定に用いられる特徴で, θ 帯域の DASM は以下の式で算出される [4].

$$\text{DASM}_\theta^{C_1, C_2} = P_\theta^{C_1} - P_\theta^{C_2} \quad (4.5)$$

ここで, C_1 と C_2 は, 左右対称の位置に配置された電極である.

4.1.4 データ整形とセッティング

本研究では, ランキング学習においてポイントワイズ手法とペアワイズ手法の比較を行う. ポイントワイズ手法では, 収集したデータ数 N 個のデータしか扱うことができないが, ペアワイズ手法で

表 4.1 nDCG による識別手法の比較

Pairwise Methods	DRNN1	DRNN2	DRNN3	RSVM
nDCG	0.73 ± 0.13	0.81 ± 0.14	0.63 ± 0.12	0.54 ± 0.13
Pointwise Methods	線形 SVR	ANN		
nDCG	0.46 ± 0.22	0.15 ± 0.13		

は、差分をとったものを 1 つのデータとして扱うため、最大で $_N P_2$ 個のデータを扱うことができる。ランキング学習に用いるランキングは、被験者により与えられたスコアに基づいて行った。ポイントワイズではデータ数 N 個、ペアワイズでは $_N P_2$ 個のランキングラベルが得られた。また、今回の測定ではトレーニングデータとテストデータを区別していない。そのため、交差検定を用いて性能を算出した。アーチファクト除去のため取り除かれたデータ数が被験者ごとに異なるため、9 個のデータをテストデータとしその他を用いて学習を行うことを可能な限り繰り返した。

RDNN のパラメータの設定は以下のようにした。学習率 ϵ を 0.005、パラメータの更新回数を 100-1000 回、層数は 6 層、各層のユニット数は入力層と出力層を除いて前の層の半分となるように決定した。プログラムは Python を用いて作成した。RSVM, SVR は線形カーネルを用いた。コストパラメータ C は $0.1 < C < 100$ から最適なパラメータを選択した。ANN は中間層の数を 2 層とし、出力をシグモイド関数を用いて 1 から 10 の範囲とした。

4.2 結果

ランキング学習手法の有効性を normalized discounted cumulative gain (nDCG) と順位付け精度の 2 つの指標で評価する。また、RDNN のパラメータの検討を行った結果を示す。

4.2.1 normalized Discounted Cumulative Gain (nDCG)

順位付けから推薦を行う場合、上位の候補から推薦が行われる。従って、上位の候補が利用者の好みと一致している場合に高い評価値となる指標が好ましい。情報検索の分野で用いられる指標として、normalized Discounted Cumulative Gain (nDCG)[22, 23] がある。次式により nDCG は計算される。

$$\text{DCG} = \sum_{i=1}^k \frac{2^{\text{rel}_i} - 1}{\log_2(i + 1)} \quad (4.6)$$

$$\text{nDCG} = \frac{\text{DCG}}{\text{idealDCG}} \quad (4.7)$$

ここで rel_i は i 番目の提案の関連度、 k を順位付けを行う最大数、idealDCG は正しく順位付けできた場合の DCG である。DCG が idealDCG と等しくなる時、nDCG=1 となり、最も高い性能であることを示す。本研究では、 rel_i は被験者が付けた楽曲の評価値、 k は 3 とした。nDCG による評価を表 4.1 に示す。

表 4.2 Classification Accuracy [%]

Pairwise Methods	DRNN1	DRNN2	DRNN3	RSVM
Accuracy	85.62 \pm 8.33	82.53 \pm 12.45	79.21 \pm 7.98	64.80 \pm 13.40
Pointwise Methods	線形 SVR	ANN		
Accuracy	62.9 \pm 18.5	38.1 \pm 23.3		

表 4.3 DRNN1 のバッチサイズ毎の平均識別率 [%]

バッチサイズ	1	3	5	7	9
識別率	61.19 \pm 11.52	74.80 \pm 7.71	76.02 \pm 8.70	78.20 \pm 9.49	77.83 \pm 9.85

表 4.4 DRNN2 のバッチサイズ毎の平均識別率 [%]

バッチサイズ	1	3	5	7	9
識別率	69.00 \pm 9.26	74.11 \pm 8.47	76.49 \pm 7.41	76.63 \pm 6.81	78.10 \pm 10.00

表 4.5 DRNN3 のバッチサイズ毎の平均識別率 [%]

バッチサイズ	1	3	5	7	9
識別率	58.42 \pm 10.29	74.45 \pm 10.97	76.16 \pm 7.18	73.12 \pm 10.35	67.37 \pm 6.00

4.2.2 順位付け精度

順位付け精度は、2つのデータの順位付けがどれだけ正しく行えたかを評価する指標である。ポイントワイズ手法においても、2つのデータのスコアを比較することで精度を算出した。結果表 4.2 に示す。

4.2.3 DRNN のパラメータの検討

DRNN のパラメータについて検討を行う。検討は一人の被験者のデータを用いて行った。各 DRNN についてバッチサイズを変更した結果を、次の図 4.2 から図 4.10 に示す。移動平均は 30 回の更新の移動平均とした。また、バッチサイズを変更した際の平均識別率の変化を次の表 4.3 から表 4.5 に示す。



図 4.2 DRNN1: バッチサイズ毎の訓練誤差

表 4.6 特徴量ごとの識別率の変化 [%]

特徴量	PSD	ShortTimePSD	DASM
識別率	72.23 ± 12.03	75.74 ± 10.49	68.01 ± 13.91

4.2.4 特徴量の検討

DRNN に用いる特徴量の比較を行った。結果を表 4.6 に示す。

4.3 考察

表 4.1, 4.2 より、ポイントワイズ手法によりペアワイズ手法が高い性能を示した。ペアワイズ手法はペアを作成することでデータ数を増やすことができ、脳波のような少数のデータから学習を行う際に有効であることがわかる。特に ANN は線形 SVR よりも低い値を示したことから、100 個以下のデータで特徴が 200 を超える次元の場合は十分に学習が行えていないことが示された。一方、ペア

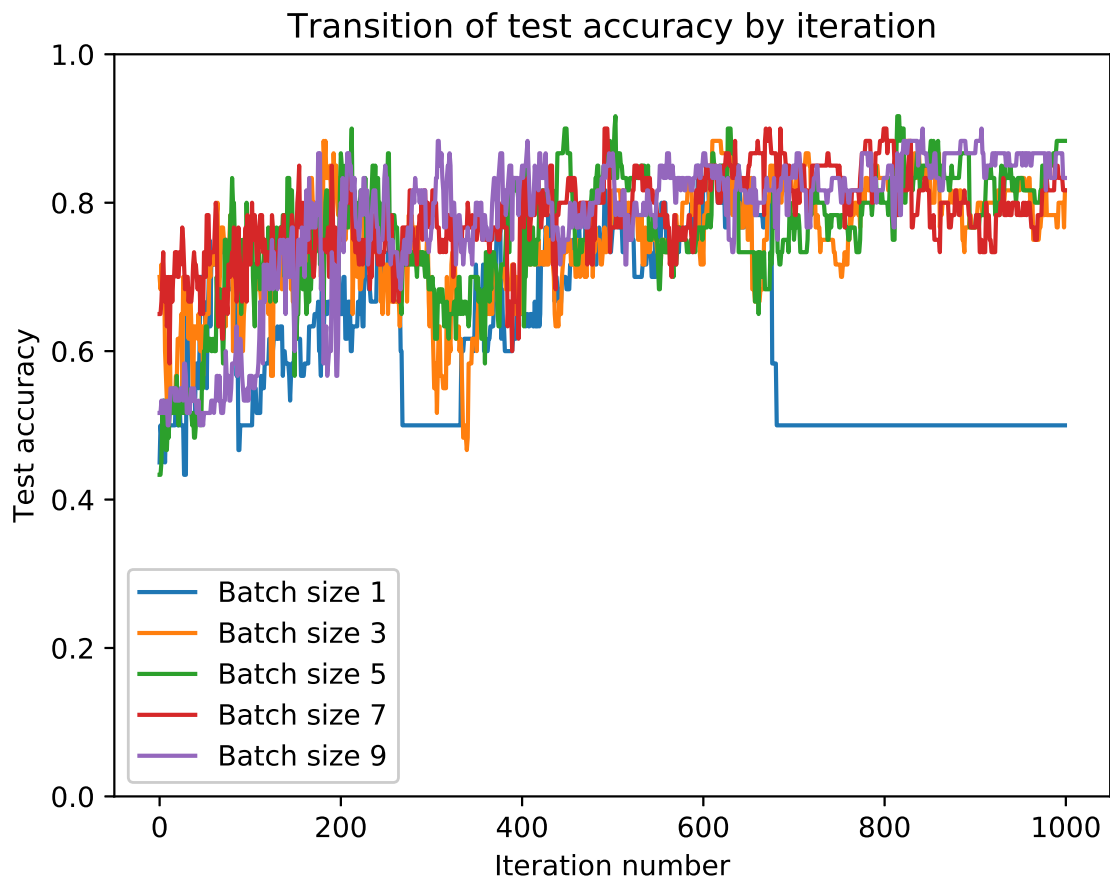


図 4.3 DRNN1: バッチサイズごとのテスト精度

ワイズ手法では少なくとも 4000 個のデータを作ることができ、RSVM と比較し DRNN1,2,3 は高い性能を示した。

表 4.1 より、nDCG は提案した DRNN の中で DRNN2 が 0.81 と最大の値を示した。DRNN3 よりも高い値を示したのは、重みにかけた制約により、不適切な局所解に収束することを避けられたことが原因であると考えられる。一方、より厳しい制約をかけた DRNN1 は DRNN2 よりも小さい値を示した。第 2 層目の制約によって下がった表現能力により、高い関連度を持つデータに対しての学習能力で劣ったと考えられる。

表 4.2 より、識別精度においては DRNN1 が最も高い値を示した。nDCG と同様に制約により不適切な局所解への収束が避けられたと考えられる。また、DRNN2 よりも高い精度を示したことから、2 層目の重みの制約も順位付けの学習においては有効であると考えられる。

また、 k の値を増加させることで、DRNN1 が DRNN2 よりも nDCG の値が高くなった。これは識別率が DRNN1 の方が高いことからわかる。本研究においては nDCG の計算の k は 3 であるが、テストデータの個数が 9 であることから、上位 33% である。音楽を推薦する状況において 100 曲の音楽があったとして、33 曲も推薦されても利用者は全てを聴くことは難しいと考えられる。よって k を増加させると結果が逆転するとしても、DRNN2 の方が提案した音楽推薦システムにおいては適当

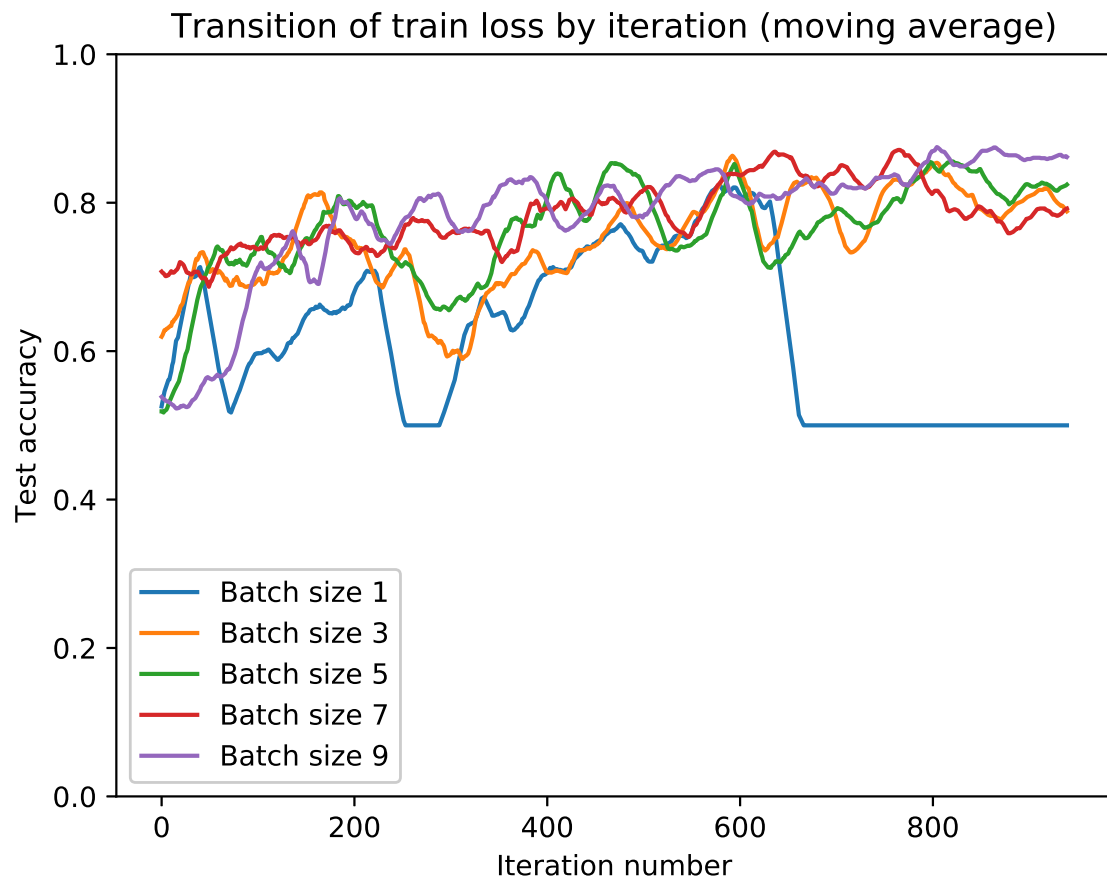


図 4.4 DRNN1: バッチサイズごとのテスト精度の移動平均

だと言える。

続いて、各 RDNN について検討を行う。まず図 4.2, 4.3, 4.4 と表 4.3 より、RDNN1 ではバッチサイズが 1 だと学習が十分に行われていないことが確認できる。またバッチサイズを増やすと、トレーニングデータに対する学習が早くなるが、サイズが 9 を超えると、テストデータに対する性能が過学習により失われることが示された。バッチサイズが大きいとテストデータに対する性能の上下が小さく、安定していることが確認出来る。以上から、RDNN1 においてはバッチサイズ 7 が最適なパラメータといえる。

図 4.5, 4.6, 4.7 と表 4.4 より、RDNN2 では、バッチサイズ 9 のとき、最も安定しており、精度も最も高かった。963 回目の更新時に、今回の識別で最も高い 98.57% の識別率が得られた。

図 4.8, 4.9, 4.10 と表 4.5 より、RDNN3 では、バッチサイズが 1 のときは他の RDNN に比べて学習が行えていないことがわかる。これは、NN の制約が少ないにも関わらず 1 つのサンプルで更新を行ったことが原因と考えられる。バッチサイズが 5 のときに最大の識別率を示し、バッチサイズが 5 から離れるにつれて識別率が低下した。また、バッチサイズが 9 のときに訓練誤差が他のバッチサイズに比べて高い値で収束していることが確認出来る。これは、不適切な局所解に収束してしまったことを示しており、バッチサイズが大きいことと、RDNN3 の表現能力が高かったことが原因と考え



図 4.5 DRNN2: バッチサイズ毎の訓練誤差

られる。

推薦に用いる特徴量については、表 4.6 より 2 秒ごとの PSD をとった特徴 (ShortTimePSD) が最も高い識別率を示した。この特徴は他の特徴を線形和で表現できるため、最も高い識別率を示したと考えられる。他の特徴と組み合わせることで必要な表現を簡単に実現できることから、組み合わせることでさらに高い識別率が示されたと考えられる。

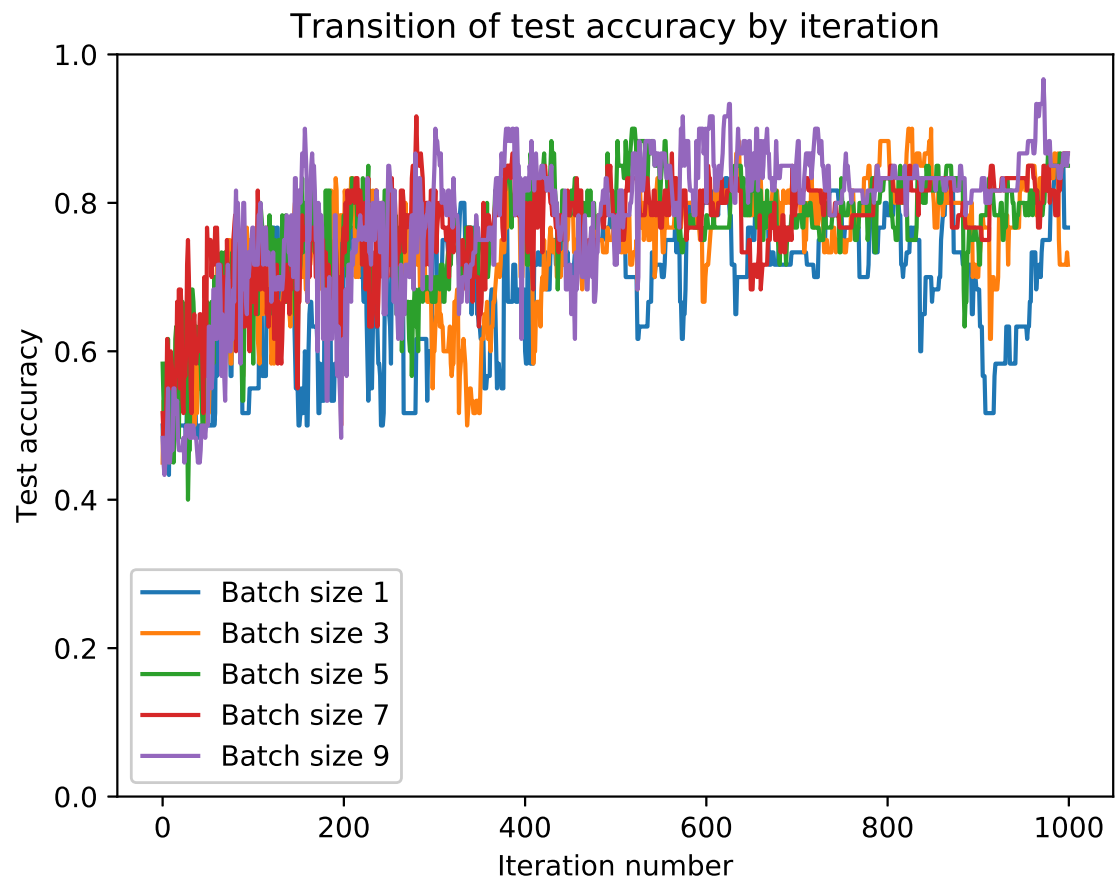


図 4.6 DRNN2: バッチサイズごとのテスト精度

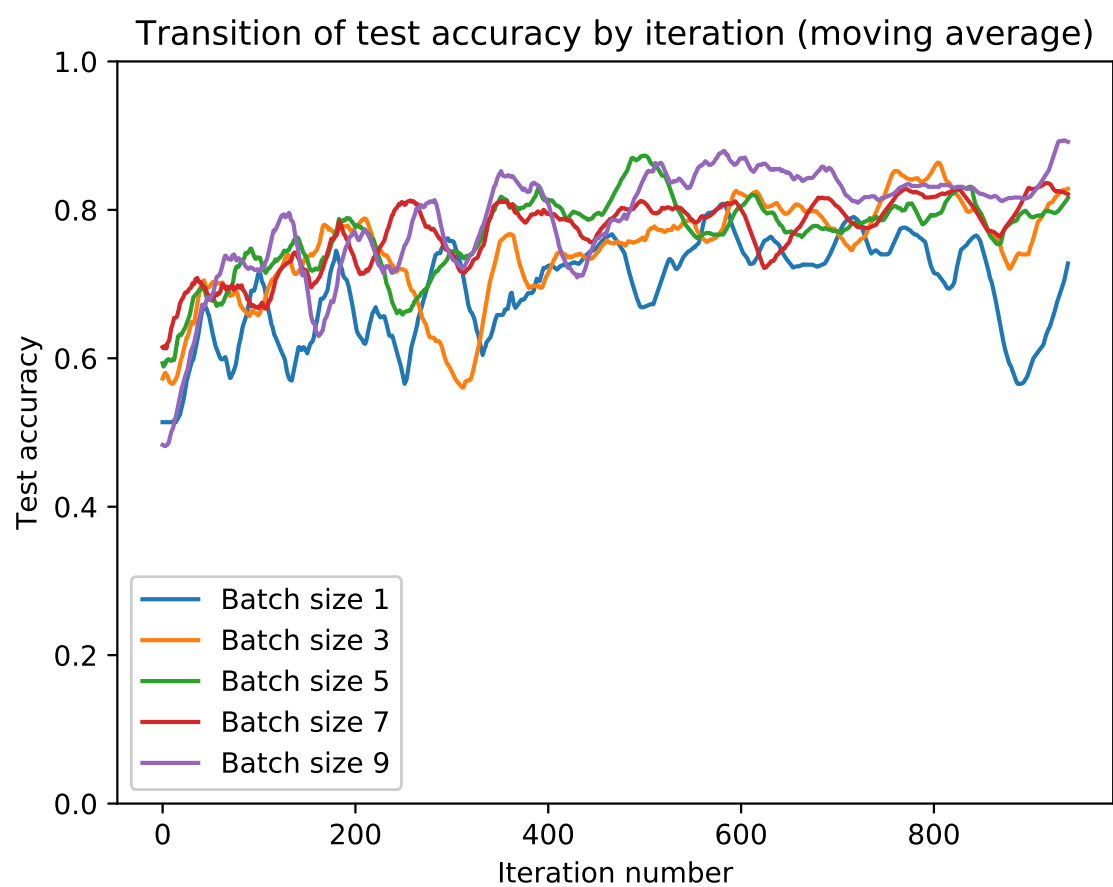


図 4.7 DRNN2: バッチサイズごとのテスト精度の移動平均

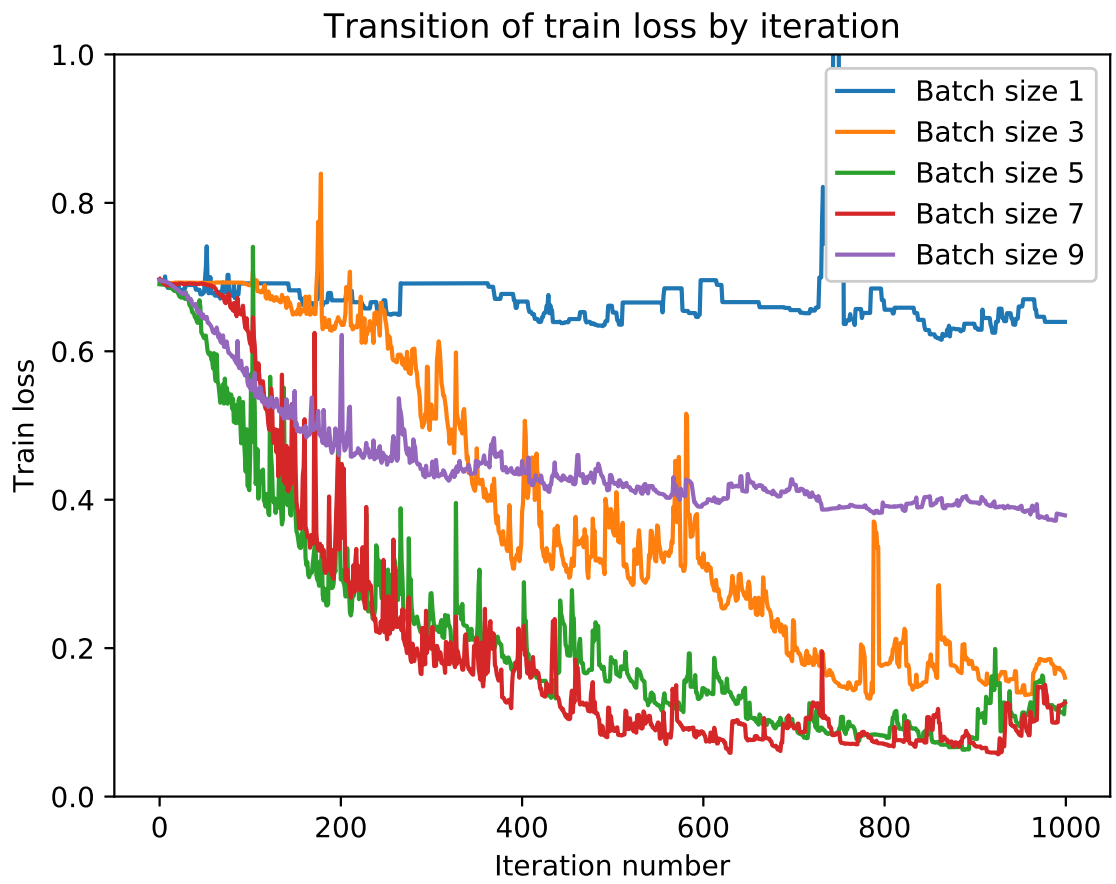


図 4.8 DRNN3: バッチサイズ毎の訓練誤差

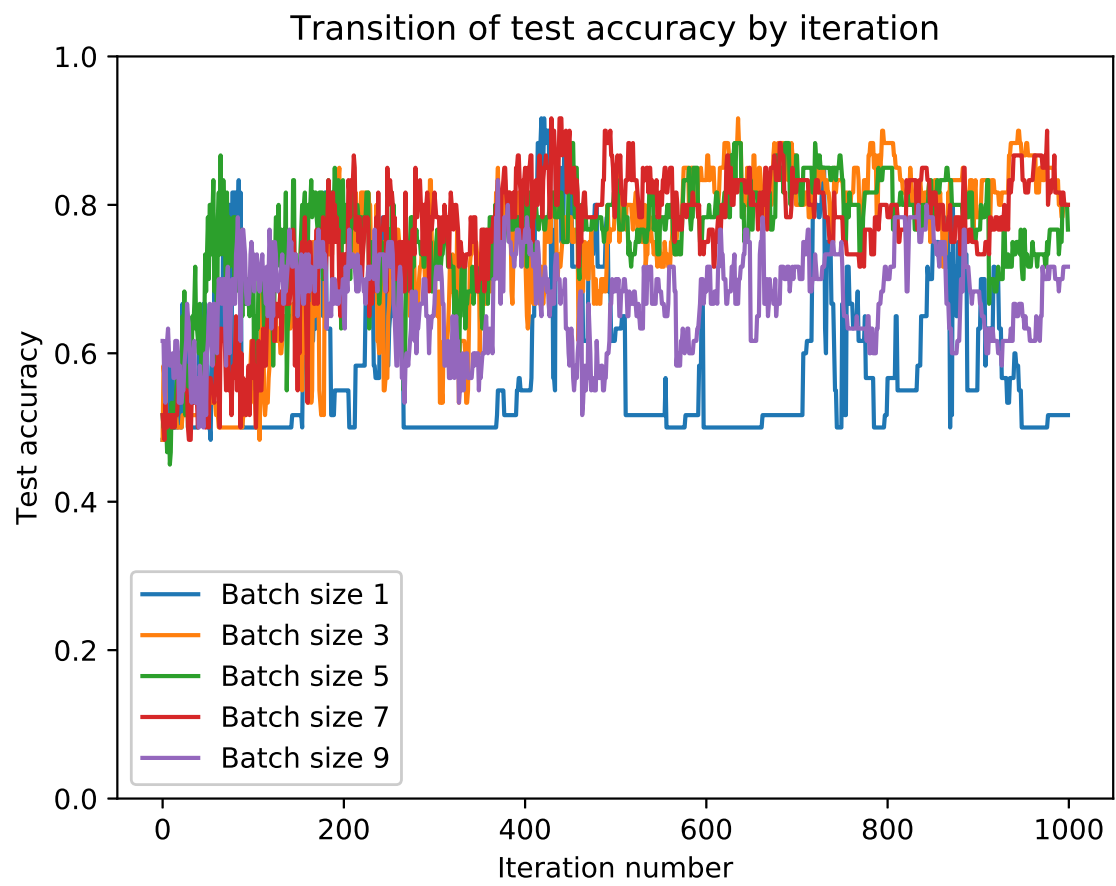


図 4.9 DRNN3: バッチサイズごとのテスト精度

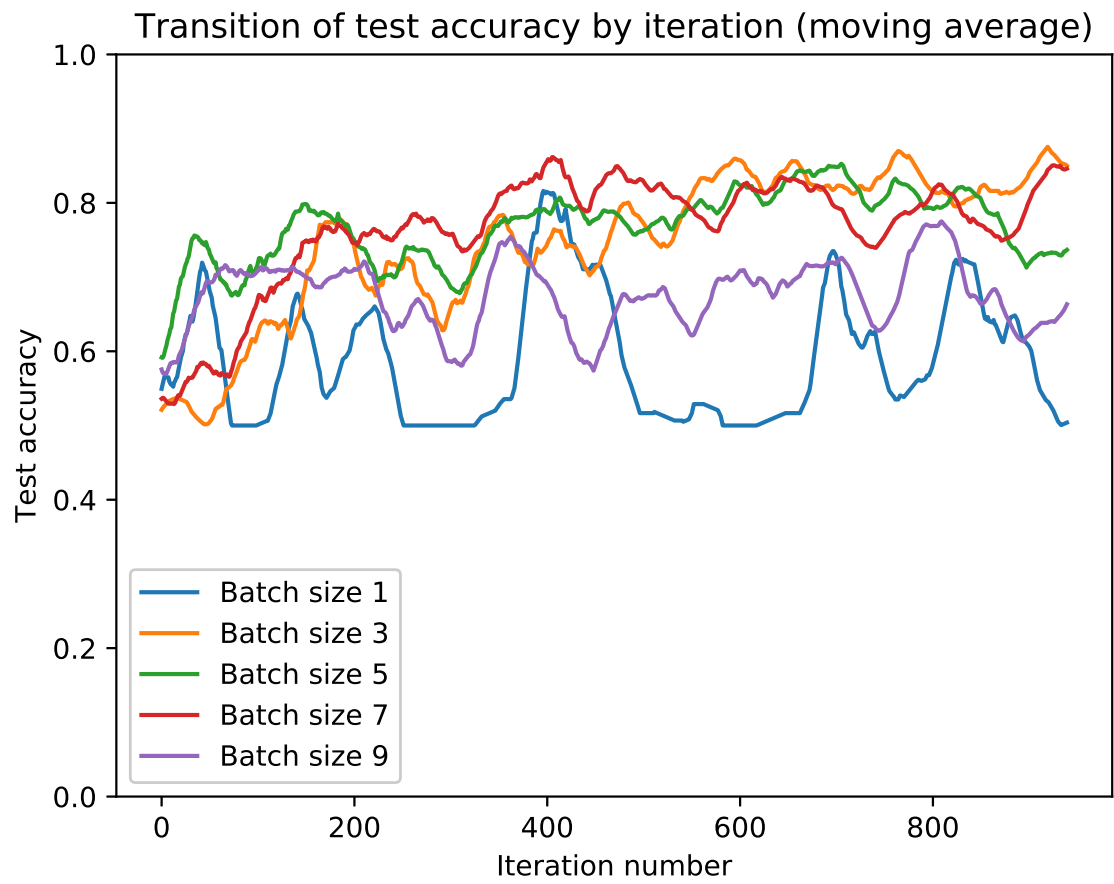


図 4.10 DRNN3: バッチサイズごとのテスト精度の移動平均

第 5 章

結論

5.1 まとめと今後の課題

本研究では，音楽鑑賞時の脳波を測定することで，利用者の好みと一致する音楽を推薦するシステムを提案した．性能評価のための実験を行った結果，推薦に用いる機械学習手法として提案した RDNN2 が最も高い性能であることが示された．RDNN2 は識別精度，nDCG の 2 つの指標で高い値を示したことから，提案システムを実際に用いることが可能であると考えられる．

今回提案した音楽推薦システムは一度聞いた音楽から好みである音楽を予測し，推薦するシステムである．一般に利用される推薦システムにおいて，利用者に新しいアイテムを推薦することは重要である．よって，協調フィルタリングなどの既存の推薦システムと組み合わせることが必要となる．

謝辞

本論文の執筆及び，研究を行うにあたりご指導してくださった鷲沢准教授に心より感謝いたします．

参考文献

- [1] G.E. Hinton and R.R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol.313, no.5786, pp.504–507, 2006.
- [2] S. Hitoshi and W. Liu, “Person authentication by multiple biometrics,” *CSEC*, vol.1999, no.45 (1999-CSEC-005), pp.37–42, 1999.
- [3] H. Tagaito, “An evaluation of user interface with emotions,” *Unisys Technology Review*, vol.64, pp.605–616, 2000.
- [4] Y.P. Lin, C.H. Wang, T.P. Jung, T.L. Wu, S.K. Jeng, J.R. Duann, and J.H. Chen, “EEG-based emotion recognition in music listening,” *IEEE Transactions on Biomedical Engineering*, vol.57, no.7, pp.1798–1806, 2010.
- [5] E. Yokomatsu, S. Ito, Y. Mitsukura, J. Cao, and M. Fukumi, “A design of the preference acquisition detection system,” *SICE, 2007 Annual Conference*, pp.2804–2807, 2007.
- [6] 糸賀弘樹, 鷲沢嘉一, “ニューラルネットワークを用いた自発脳活動解析による音楽ジャンルの推定 (応用音響),” *電子情報通信学会技術研究報告*, vol.116, no.475, pp.119–122, 2017.
- [7] T. Joachims, “Optimizing search engines using clickthrough data,” *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.133–142, 2002.
- [8] 武田常広, *脳工学*, コロナ社, 2003.
- [9] G.H. Klem, H.O. Lüders, H. Jasper, and C. Elger, “The ten-twenty electrode system of the international federation,” *Electroencephalography and Clinical Neurophysiology*, vol.52, p.3, 1999.
- [10] 時実利彦, 藤森聞一, 島蘭安雄, 佐野圭司, *新脳波入門*, 南山堂, 1969.
- [11] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, “Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music,” *Psychophysiology*, vol.44, no.2, pp.293–304, 2007.
- [12] L. Aftanas and S. Golocheikine, “Human anterior and frontal midline theta and lower alpha reflect emotionally positive state and internalized attention: high-resolution EEG investigation of meditation,” *Neuroscience Letters*, vol.310, no.1, pp.57–60, 2001.
- [13] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender, “Learning to rank using gradient descent,” *Proceedings of the 22nd International Conference on Machine Learning*, pp.89–96, 2005.

- [14] T.-Y. Liu, et al., “Learning to rank for information retrieval,” *Foundations and Trends in Information Retrieval*, vol.3, no.3, pp.225–331, 2009.
- [15] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li, “Learning to rank: from pairwise approach to listwise approach,” *Proceedings of the 24th International Conference on Machine Learning*, pp.129–136, 2007.
- [16] 竹内一郎, 烏山昌幸, 機械学習プロフェッショナルシリーズ サポートベクトルマシン, 講談社, 2015.
- [17] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer science & business media, 2013.
- [18] M. Welling, “Support vector regression,” *Department of Computer Science, University of Toronto*, Toronto, 2004.
- [19] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp.315–323, 2011.
- [20] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, pp.1097–1105, 2012.
- [21] A.V.D. Oord, S. Dieleman, and B. Schrauwen, “Deep content-based music recommendation,” *Advances in Neural Information Processing Systems 26*, eds. by C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, pp.2643–2651, Curran Associates, Inc., 2013.
- [22] K. Järvelin and J. Kekäläinen, “IR evaluation methods for retrieving highly relevant documents,” *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.41–48, 2000.
- [23] K. Järvelin and J. Kekäläinen, “Cumulated gain-based evaluation of ir techniques,” *ACM Transactions on Information Systems (TOIS)*, vol.20, no.4, pp.422–446, 2002.