

平成 29 年度 卒業研究論文

題目 台詞の重回帰分析によるアニメ劇伴の
BPM 推定に関する研究

指導教員 服部 峻

提出者 室蘭工業大学 情報電子工学系学科
氏名 伊藤 優太
学籍番号 14024021

提出年月日 平成 30 年 2 月 13 日

目次

第 1 章	まえがき	1
第 2 章	劇伴制作の流れ	2
第 3 章	提案システム	4
3.1	劇伴自動作曲システムの概要	4
3.2	BPM 推定手法の概要	5
第 4 章	劇伴の BPM に影響を与える要因の検討	8
4.1	アニメ映像中の動作と劇伴の BPM の相関分析	8
4.2	BPM に影響を与える要因の検討結果に基づく予測式	10
4.3	予備実験	10
第 5 章	提案手法	12
5.1	SF-ISF による特徴語抽出と辞書作成	13
5.2	台詞と辞書を用いた重回帰分析による予測式	15
第 6 章	評価実験	16
第 7 章	まとめと今後の課題	21
	謝辞	23
	参考文献	24

目次

2.1	会社制作のアニメにおける劇伴制作の流れ	3
3.1	劇伴の自動作曲システム（個人制作アニメ）	6
3.2	劇伴の自動作曲システム（会社制作アニメ）	6
3.3	BPM 推定手法（個人制作アニメ）	7
3.4	BPM 推定手法（会社制作アニメ）	7
4.1	場面の定義	8
4.2	「走っている場面か否か」と BPM の関係	9
4.3	「戦っている場面か否か」と BPM の関係	9
4.4	ギャグアニメとバトルアニメの合格率の比較	11
5.1	提案手法の流れ	12
6.1	辞書セット 1（シソーラスを基に手製した辞書）	16
6.2	辞書セット 2（SF-ISF による特徴語辞書を含む）	17
6.3	辞書セット 1（シソーラスを基に手製した辞書）の MSE	17
6.4	辞書セット 2（SF-ISF による特徴語辞書を含む）の MSE	18
6.5	2 種類の辞書セットにおける推定 BPM の合格率	18
6.6	MSE 及び合格率（閾値 0.00~0.75）	19
6.7	MSE 及び合格率（閾値 1.00~1.75）	20
6.8	MSE 及び合格率（閾値 2.00~3.00）	20
6.9	閾値と MSE の関係	20

表目次

5.1	「走っている場面」(左)と「戦っている場面」(右)の特徴語 (SF_p のみ)	14
5.2	「走っている場面」(左)と「戦っている場面」(右)の特徴語 (SF-ISF)	14

第 1 章

まえがき

劇伴とはテレビドラマ，アニメ，映画，演劇などの劇中音楽のことであり，それぞれの場面の印象を決定づける重要な要素で，人間の手で作られている．プロが劇伴を作る場合，作曲家や音楽プロデューサー，監督や音響監督などの多くの人が関わっている．また，劇伴の制作には 1 クールのアニメで 45～60 日程度と大きなコストが掛かっている．一方，個人制作のアニメの場合，制作者は絵を描くスキルを持っていても，音楽に関するスキルが無く自分で劇伴を作れないことも多い．自分で劇伴を作ることができない場合，劇伴として既存のゲームやアニメの曲を使うこともあるが，この場合制作したアニメを公開すれば，著作権を侵害することになるので，自由に公開できない．

そこで本研究では，これらの問題を解決するために，アニメの場面に合った劇伴を自動で作曲するシステムを提案する．その第一歩として，劇伴の構成要素の 1 つである BPM (Beats Per Minute) を推定する手法について検討する．まず，個人制作と会社制作のアニメに分けて，劇伴制作の流れについて整理した上で，それぞれに対する BPM 推定手法を提案する．本稿の主な対象である個人制作のアニメ映像の BPM を推定する場合，あらかじめアニメ映像から台詞と劇伴の BPM を取得して重回帰分析を行い，BPM の予測式を求めておく．その後，劇伴を付けたい場面の台詞を取得し，求めておいた予測式を用いて劇伴の適切な BPM を推定する．

自動作曲に関する関連研究としては，歌詞を入力とした自動作曲 [1] や既存の楽曲データを入力とした自動作曲 [2] などが挙げられるが，劇伴などの場面に合わせた曲の自動作曲に関する研究はあまり行われていない．

第 2 章

劇伴制作の流れ

本章では、個人制作アニメと会社制作アニメに分けて、それぞれにおける劇伴制作の流れについて整理する [3,4]。まず、個人制作アニメにおける劇伴は、既存の曲を劇伴として使用することも多いが、アニメ映像の制作者が制作する場合は 1 人で作ることになる。そのため、会社制作アニメの劇伴制作における音楽メニューなどのドキュメントが作られないことが多く、アニメ映像を基にして劇伴を制作する。

一方、図 2.1 のように、会社制作アニメの劇伴制作は集団で行われ、意思疎通が必要であるため、ドキュメントが用いられる。作業 1 では、音楽プロデューサーが監督に作曲家を提案する。まず、監督が音楽プロデューサーに「こんな世界観でいきたい」などといった抽象的なオーダーを出す。音楽プロデューサーはそのオーダーに対して、監督との相性やスケジュールを考慮し、どの作曲家が適しているかを考えて作曲家を提案する。作業 2 では、音響監督が「音楽メニュー」と呼ばれる楽曲のリストを作成する。音響監督と監督が話し合い、そこで音響監督がシナリオやコンテを読み解いていく。その後、音響監督が話し合いの内容を基に音楽メニューを作成する。音楽メニューには、曲番号や仮タイトル、音響監督のコメントなどが書かれている。作業 3 では、「劇伴メニュー打ち合わせ」が行われる。音響監督から劇伴に関する具体的なオーダーが示される。例えば、「このアニメは 40 曲必要」「それぞれの劇伴の尺はこれくらい」といったオーダーが示される。監督からは、今回のアニメで描きたいことなど、全体的な方向性が示される。これらのオーダーを音楽プロデューサーと作曲家は理解する。作業 4 では、音楽メニューを基にして劇伴の作曲を行う。音楽プロデューサーがスケジュール管理、予算管理を行い、作曲家はスケジュールに合わせて劇伴を作っていく。作曲家がデモ音源を作り、それを音楽プロデューサーがチェックし、全体の方向などの指示を出していく。作業 5 では、音響監督が映像に劇伴を割り当てる。この作業に音楽プロデューサーと作曲家は関わらないが、音響監督は演出の仕方が分かっているので良い劇伴になる。

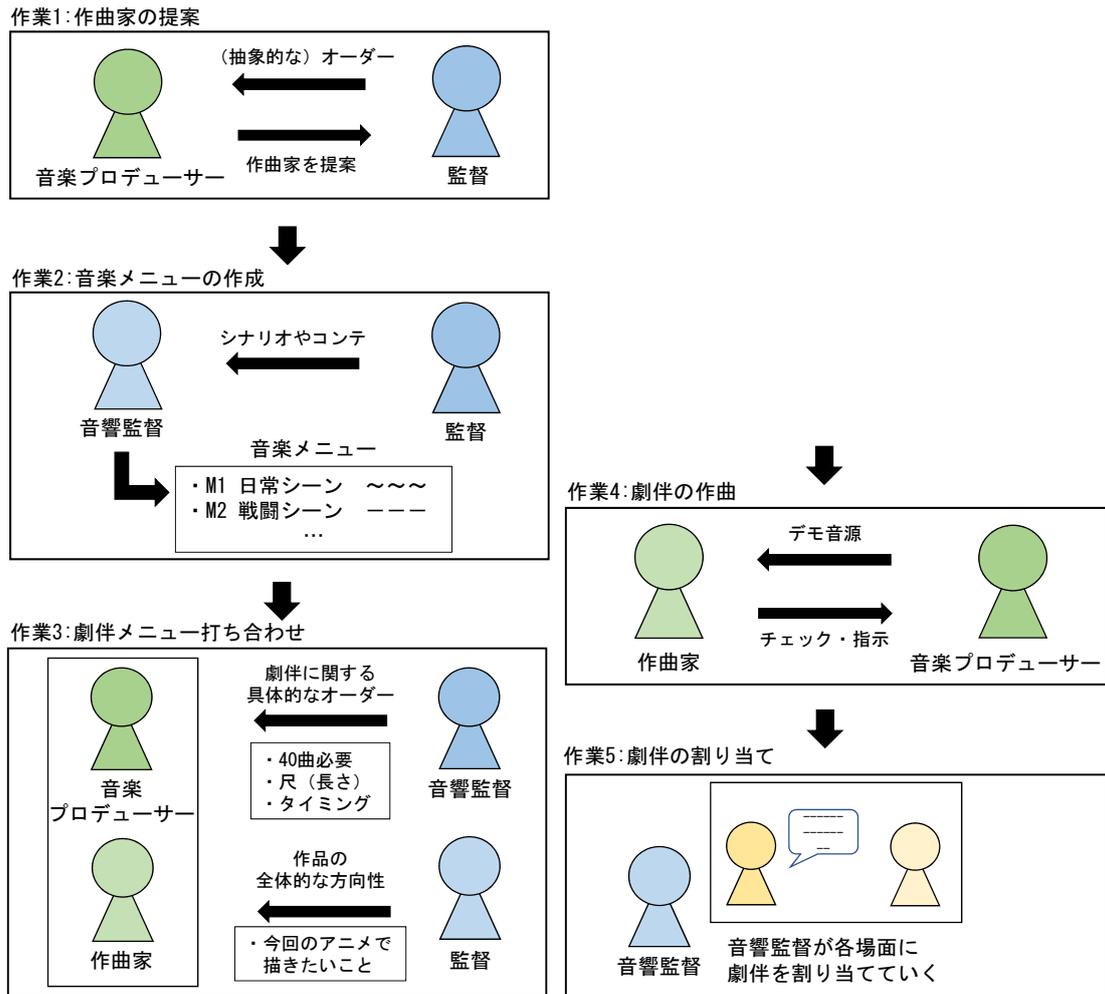


図 2.1 会社制作のアニメにおける劇伴制作の流れ

第3章

提案システム

本章では、劇伴自動作曲システムの概要、及び、劇伴の構成要素の1つであるBPMを推定する手法の概要について述べる。

3.1 劇伴自動作曲システムの概要

本稿では、個人制作のアニメの劇伴を作る場合と、会社で制作しているアニメの劇伴を作る場合の2つの場合を想定する。提案するシステムのうち、個人制作アニメの劇伴の自動作曲システムの処理を図3.1に、会社制作アニメの劇伴の自動作曲システムの処理を図3.2に示す。会社で制作しているアニメの場合、劇伴の制作は「音楽メニュー」と呼ばれる楽曲のリストを基にして行われているので、システムへの入力には音楽メニューとする。一方、個人制作のアニメの場合はアニメ映像を作った後に劇伴を作る作業を行うことができ、音楽メニューを使わないのでシステムへの入力にはアニメ映像とする。

個人制作のアニメの劇伴を作る場合、処理1では台詞などのテキスト情報、輝度や輝度変化などの画像情報、登場人物の声のスピードや抑揚などの音声情報を取得する。処理2では処理1で取得した情報を用いてどのような場面であるか判定した上で数値化する。処理3では処理2で判定した場面に合う劇伴を作曲する。処理4では完成した曲を音声ファイルで出力する。

会社で制作しているアニメの劇伴を作る場合、処理1では音楽メニューからテキスト情報を取得する。テキストには、仮タイトルや場面の種類（「日常シーン」「戦闘シーン」など）、音響監督や音楽プロデューサーからのコメントが含まれる。例えば、その場面で伝えたいことや、音楽プロデューサーなりの場面の解釈がコメントとして書かれている。処理2では処理1で取得したテキストを用いてどのような場面であるか判定する。処理3では処理2で判断した場面に合う劇伴を作曲する。処理4では完成した曲を音声ファイルで出力する。

ここでは2つの場合を想定したが、劇伴の作曲が自動化できた場合、個人制作のアニメの劇伴の方が恩恵が大きい。個人制作のアニメの場合、会社で制作している場合と違い、劇伴の作り方自体が分からず、自分が作ったアニメに劇伴を付けることができないという大きな問題があるので、本稿では個人制作のアニメの劇伴を主な対象とする。

劇伴の構成要素としてはメロディ、コード進行、BPMなどが挙げられるが、本稿ではその1

つである BPM に着目し、アニメ映像を入力として用いた場合の BPM 推定手法を提案する。

3.2 BPM 推定手法の概要

個人で制作しているアニメの劇伴を作る場合の BPM 推定手法の概要を図 3.3 に、会社で制作しているアニメの劇伴を作る場合の BPM 推定手法の概要を図 3.4 に示す。BPM 推定手法は、サンプルデータから予測式を求める「学習フェーズ」と、求めた予測式から BPM を推定する「推定フェーズ」から成る。

個人制作アニメの劇伴を作る場合、処理 1 ではアニメ映像から台詞と劇伴の BPM を取得する。処理 2 では取得した台詞と BPM を用いて重回帰分析を行い、BPM の予測式を求める。重回帰分析の目的変数 y は BPM であり、説明変数 x_i は BPM に影響を与える要因について表した値である。例えば、ある場面に出て来る登場人物が走っているか否かを表す変数などを用いる。処理 3 では劇伴を付けたい場面の台詞を取得し、処理 2 で求めた予測式を用いて劇伴の適切な BPM を推定する。

会社制作アニメの劇伴を作る場合、処理 1 では音楽メニューからテキスト情報（仮タイトルや音響監督のコメントなど）と、音楽メニューを基にして作られた劇伴から BPM を取得する。処理 2 では取得したテキストと BPM を用いて重回帰分析を行い、BPM の予測式を求める。個人制作アニメの劇伴を作る場合と同様に、重回帰分析の目的変数 y は BPM であり、説明変数 x_i は BPM に影響を与える要因について表した値である。処理 3 では劇伴を付けたい場面の音楽メニューを取得し、処理 2 で求めた予測式を用いて劇伴の適切な BPM を推定する。

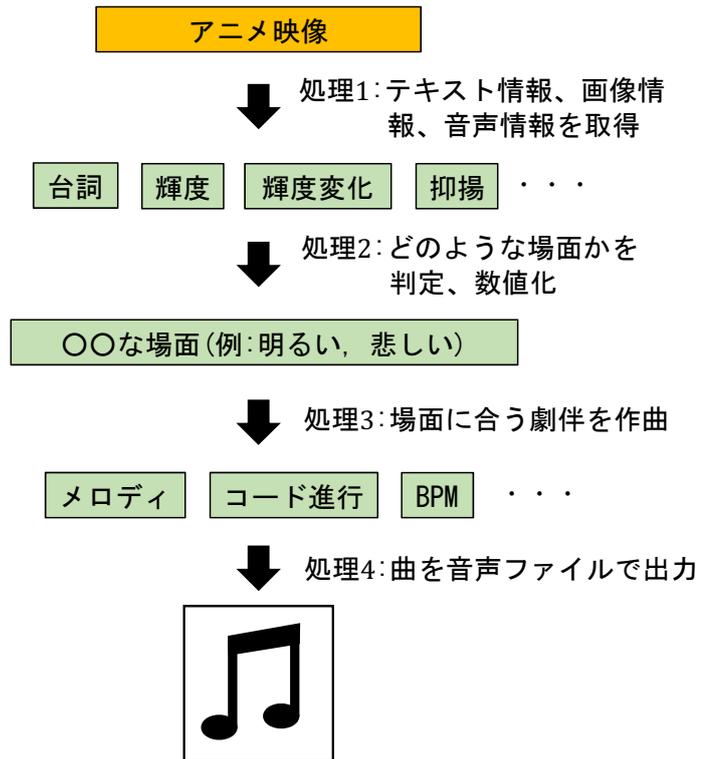


図 3.1 劇伴の自動作曲システム（個人制作アニメ）

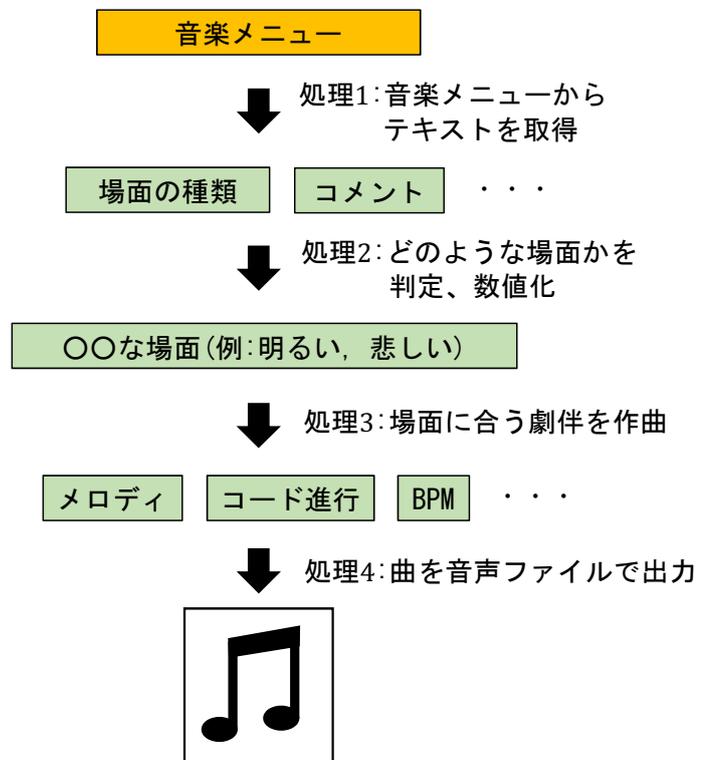


図 3.2 劇伴の自動作曲システム（会社制作アニメ）

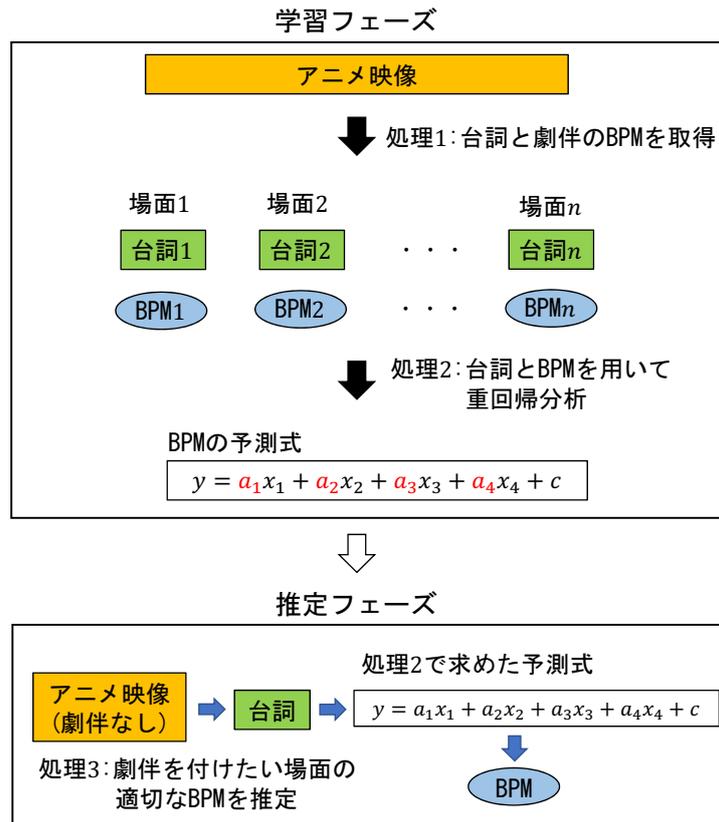


図 3.3 BPM 推定手法 (個人制作アニメ)

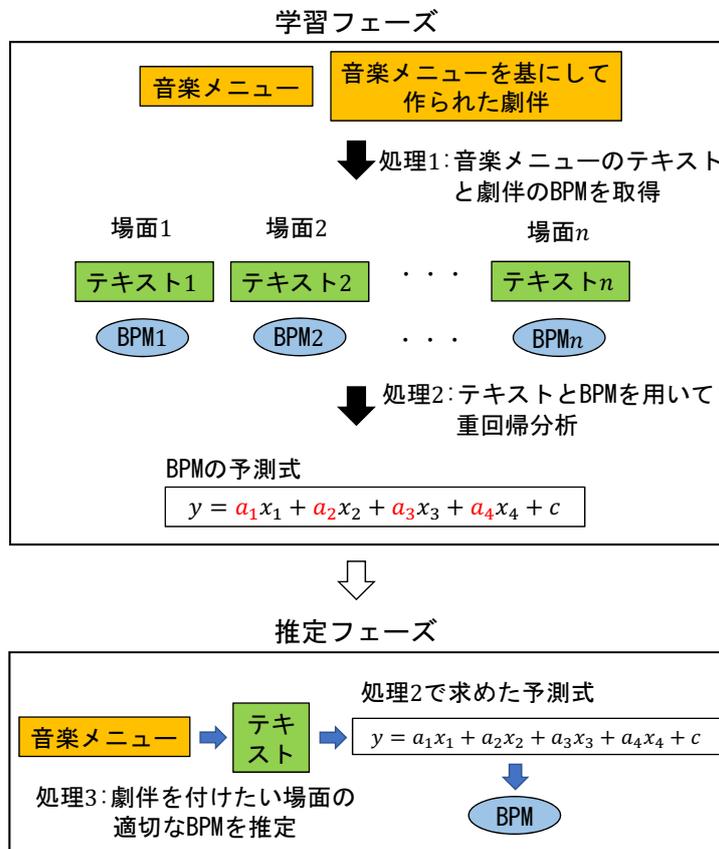


図 3.4 BPM 推定手法 (会社制作アニメ)

第4章

劇伴の BPM に影響を与える要因の検討

劇伴の BPM には、登場キャラクターの感情や動作に関する内容が影響すると考えられる。本章では、BPM 推定の手掛かりを得るために、アニメのある場面が「走っている場面か否か」「戦っている場面か否か」と BPM の関係性を調査する。

4.1 アニメ映像中の動作と劇伴の BPM の相関分析

「走っている場面か否か」や「戦っている場面か否か」と BPM の関係を散布図で分析する。説明変数は「走っている場面か否か」「戦っている場面か否か」を表す二値、目的変数は BPM の整数値である。分析に用いるサンプルデータは場面ごとに作成する。場面の定義を図 4.1 に示す。場面の定義は、ある 1 つの劇伴が始まってから終わるまでの間を 1 つの場面とする。「走っている場面か否か」については、ある場面の登場人物が一人でも走っていれば、その場面は「走っている場面」とし、値を 1 とする。一方、ある場面の登場人物が誰も走っていない場合、その場面は「走っていない場面」とし、値を 0 とする。「戦っている場面か否か」については、ある場面の登場人物が戦っていれば、その場面は「戦っている場面」とし、値を 1 とする。一方、ある場面の登場人物が誰も戦っていない場合、その場面は「戦っていない場面」とし、値を 0 とする。BPM は実際に劇伴を聴いて手動で計測した値を小数点以下で四捨五入し、4.3 節の予備実験で正解 BPM として用いる。



図 4.1 場面の定義

「走っている場面か否か」と BPM の関係、「戦っている場面か否か」と BPM の関係について、それぞれ「バトル」「ギャグ」の2種類のジャンルのアニメにおいて分析すると、「走っている場面か否か」と BPM の関係を表す散布図は図 4.2 のように、「戦っているか否か」と BPM の関係を表す散布図は図 4.3 のようになる。図 4.2 から、ギャグアニメ、バトルアニメ共に走っている場面の方が走っていない場面より BPM が高い傾向があることが分かる。図 4.3 から、ギャグアニメ、バトルアニメ共に戦っている場面の方が戦っていない場面より BPM が高い傾向があることが分かる。また、図 4.3 のギャグアニメとバトルアニメを比較すると、ギャグアニメの BPM はばらつきが大きい、バトルアニメの BPM はばらつきが少なく、高いものが多いことが分かるが、バトルアニメの戦っている場面は決着を付けるなどの大きな意味があり、BPM が高く激しい劇伴が付けられることが多いからである。

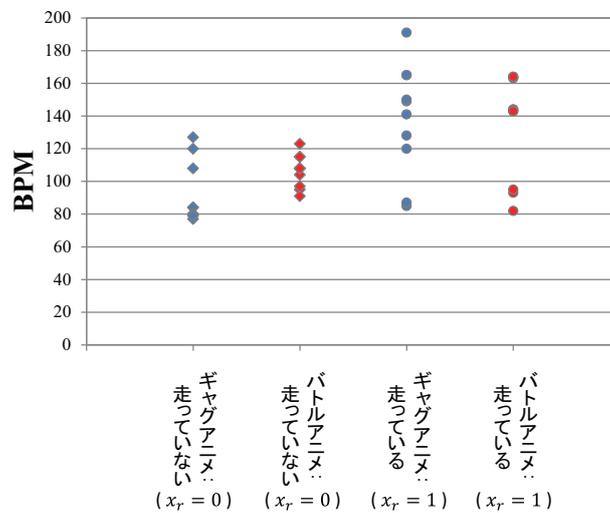


図 4.2 「走っている場面か否か」と BPM の関係

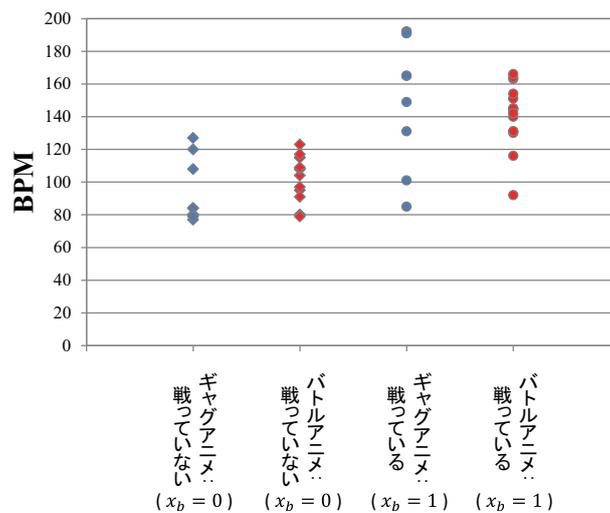


図 4.3 「戦っている場面か否か」と BPM の関係

4.2 BPM に影響を与える要因の検討結果に基づく予測式

BPM に影響を与える要因について前節で検討したが、その結果に基づく予測式を以下の式 (4.1) に示す。「走っている場面か否か」と「戦っている場面か否か」は BPM に影響することが分かったので、予測式の説明変数 x_i に「走っている場面か否か」と「戦っている場面か否か」を表す値を用い、目的変数 y に BPM を用いる。「走っている場面か否か」を表す変数を x_r ($\in \{0, 1\}$)、「戦っている場面か否か」を表す変数を x_b ($\in \{0, 1\}$) とし、それぞれの係数を a_r , a_b とする。 c は定数項である。 x_r , x_b の定義は 4.1 節と同様である。

$$y = a_r x_r + a_b x_b + c \quad (4.1)$$

4.3 予備実験

4.2 節で定義した予測式 (4.1) による BPM の推定精度を確かめる実験を行う。初めにサンプルデータを入力して予測式を求め、その後、求めた予測式を使ってサンプルデータの説明変数から BPM を推定し、正解 BPM との誤差を求める。その誤差が許容誤差以下なら「合格」、許容誤差より大きいなら「不合格」とし、許容誤差を変化させながら合格率を求める。合格率は「合格」のデータ数をデータの総数で割ったものである。

ギャグアニメとバトルアニメにおける許容誤差と合格率の関係及び平均二乗誤差 (MSE) を図 4.4 に示す。文献 [4] によると、BPM85~100 は通常の会話、BPM100~120 は歩くような動きのある軽快なテンポというように、場面ごとの BPM の範囲が 15~20 程度となっており、誤差が 10 未満であればおおよそ BPM が推定できていると言える。許容誤差が 10 前後のときの合格率は、ギャグアニメでは 0.4 程度、バトルアニメでは 0.2 程度となっており、あまり高い値とはならなかったため、「走っているか否か」「戦っているか否か」という要因のみでは不十分であり、他の要因も考慮する必要があると考えられる。2つのジャンルを比較すると、バトルアニメの方が平均二乗誤差が小さく、全体的に合格率が高いことが分かる。これは、4.1 節の分析で分かったように、バトルアニメの方がギャグアニメより「戦っている場面か否か」と BPM の相関が強いからであると考えられる。

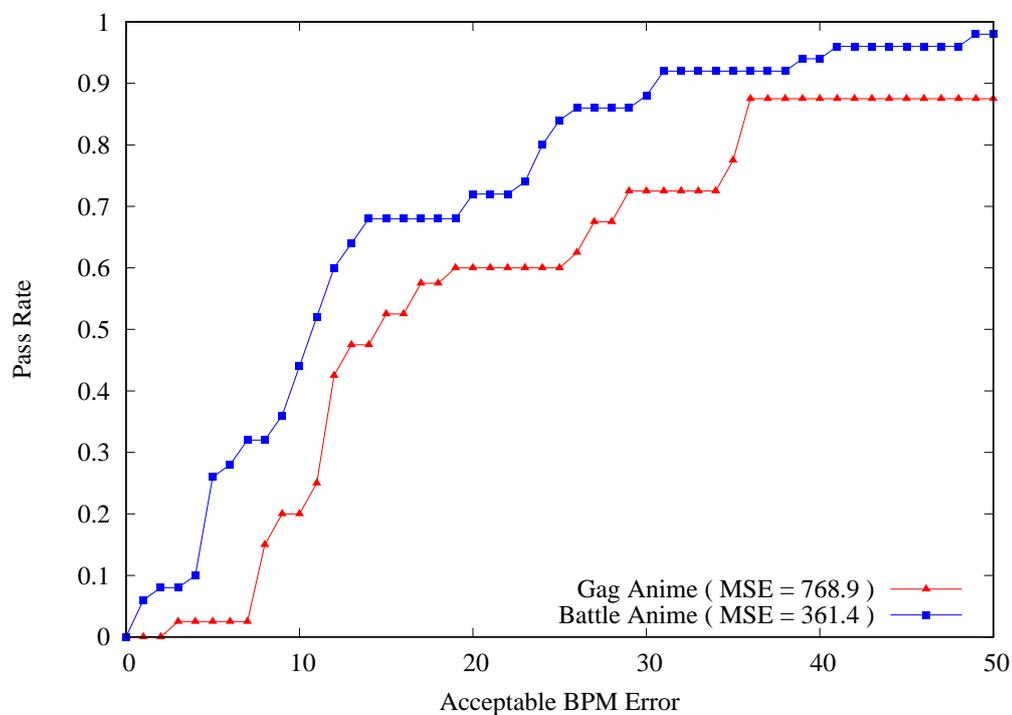


図 4.4 ギャグアニメとバトルアニメの合格率の比較

第5章

提案手法

登場人物の台詞には場面の内容を表す単語が含まれると仮説を立て、4章では手動で行った「走っている場面か否か」「戦っている場面か否か」の判断を、登場人物の台詞を用いて自動化する。提案手法の流れを図5.1に示す。まず、「走っている場面」「戦っている場面」の特徴語を登録した辞書を自動で各々作成し、日本語シソーラス連想類語辞典 [6] を用いて「喜び/楽しみ」「怒り」「悲しみ」の感情語辞書を作成する。次に、台詞と辞書を使って動作要因を数値化する計算式を定義する。最後に、これらを説明変数としてサンプルデータから予測式を求め、BPM を推定する。

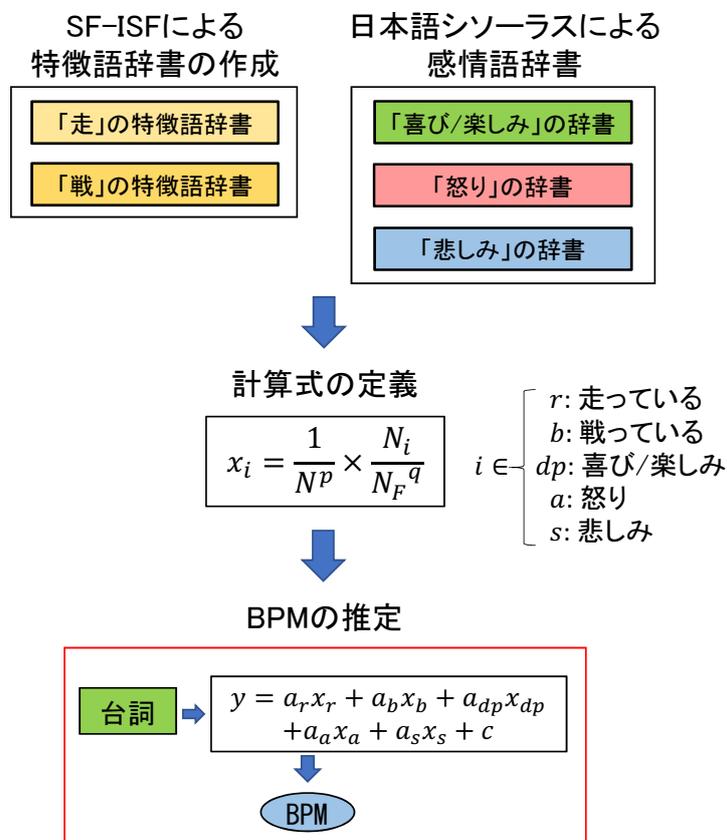


図 5.1 提案手法の流れ

5.1 SF-ISF による特徴語抽出と辞書作成

「走っている場面」「戦っている場面」の特徴語辞書を各々作成するために、TF-IDF 法を参考にした SF-ISF (Scene Frequency - Inverse Scene Frequency) による特徴語抽出を提案する。台詞中から形態素解析器 gosen (IPA 辞書版) [7] によって切り出した単語 w に対して、特徴語としての重要度を表す $SF-ISF(w)$ を以下の式 (5.1) で定義する。

$$SF-ISF(w) = SF_p(w) \times \frac{1}{SF_n(w) + 1} \quad (5.1)$$

但し、 $SF_p(w)$ は「走っている場面」「戦っている場面」の中で単語 w が含まれる場面数、 $SF_n(w)$ は「走っていない場面」「戦っていない場面」の中で単語 w が含まれる場面数である。

$SF_p(w)$ のみを用いて抽出した「走っている場面」の特徴語の上位 10 件、及び、「戦っている場面」の特徴語の上位 10 件を表 5.1 に示す。 $SF-ISF(w)$ を用いて抽出した「走っている場面」の特徴語の上位 10 件、及び、「戦っている場面」の特徴語の上位 10 件を表 5.2 に示す。本稿では、「走っている場面」の台詞データ 25 個と「走っていない場面」の台詞データ 25 個、及び、「戦っている場面」の台詞データ 25 個と「戦っていない場面」の台詞データ 25 個を用い、 $SF_p(w)$ 、及び、 $SF-ISF(w)$ が閾値 1 より大きい単語 w を各々の特徴語辞書に登録した。

単に $SF_p(w)$ の値を用いて抽出した語には句読点や記号、助詞が多く、各々の場面に関する特徴語を上手く抽出できていないのに対し、提案した $SF-ISF(w)$ の値を用いて抽出した結果には表 5.2 のように「待つ」「逃げる」「観念」「捕まえる」といった単語が含まれており、走っている場面に関する特徴語を抽出できていることが分かる。これらの単語は「待て」「逃げろ」「観念しろ」「捕まえろ」というように、命令形で走っている場面に出現することが多い。また、「よし子」「アイズ」などの呼称（キャラクター名）も含まれていることが分かる。一方、戦っている場面に関しては、「てめえ」「上手い」といった特徴語を抽出できていることが分かる。戦っている場面では「てめえ」「こいつ」などの敵対的な呼び掛けや、「上手い」「強い」などの相手を評価する単語が多く出現する。

表 5.1 「走っている場面」(左)と「戦っている場面」(右)の特徴語 (SF_pのみ)

順位	特徴語	SF _p
1	、	23
1	。	23
3	て	21
4	は	19
4	だ	19
6	!	18
6	?	18
8	の	17
9	が	16
9	する	16

順位	特徴語	SF _p
1	、	25
2	て	24
2	だ	24
4	は	21
4	の	21
4	!	21
7	た	20
7	?	20
7	。	20
7	に	20

表 5.2 「走っている場面」(左)と「戦っている場面」(右)の特徴語 (SF-ISF)

順位	特徴語	SF-ISF
1	待つ	7.00
2	逃げる	5.00
3	上げる	3.00
3	よし子	3.00
3	うわ	3.00
3	られる	3.00
7	観念	2.00
7	逃す	2.00
7	捕まえる	2.00
7	アイズ	2.00

順位	特徴語	SF-ISF
1	魔	4.00
1	られる	4.00
1	出る	4.00
4	やる	3.50
5	てめえ	3.00
5	使う	3.00
5	誰	3.00
5	秘技	3.00
5	上手い	3.00
5	逃げる	3.00

5.2 台詞と辞書を用いた重回帰分析による予測式

アニメ劇伴で使われる楽曲の種類には以下の4種類がある [5].

- 情景曲：美術を補って作品の世界観を作る曲
- 感情曲：登場人物の感情に対して付けられる曲
- 状況曲：場面の状況を説明する曲
- 状況内音楽：映像の中で実際に流れている音楽

アニメ劇伴で使う音楽には「感情曲」と呼ばれる曲があり、劇伴の制作には、「走る」「戦う」という動作だけではなく、登場人物の感情も考慮されていることが分かる。そこで本稿では、BPMに影響を与える要因として、4.1節で検討した「走る」「戦う」という動作内容に加えて、「喜/楽」「怒」「哀」という感情も採用し、それぞれの要因を表すような単語から成る辞書を用意しておく。これらの辞書とアニメ映像中の台詞を用いた重回帰分析の説明変数 x_i ($i \in \{r, b, dp, a, s\}$) を式 (5.2) で定義する。 i は説明変数の種類を表し、 r は「走っている」、 b は「戦っている」、 dp は「喜び/楽しみ」、 a は「怒り」、 s は「悲しみ」を表す。

$$x_i = \frac{1}{N^p} \times \frac{N_i}{N_F^q} \quad (5.2)$$

但し、 N はある台詞テキストが与えられたときの台詞テキスト中の全単語数、 N_i は BPM に影響を与える要因 i を表す辞書に登録されている単語の出現回数、 $N_F (= N_r + N_b + N_{dp} + N_a + N_s)$ は各々の辞書に登録されている単語の出現回数の合計である。 p, q はパラメータであり、 $p (\in \{0, 0.1, \dots, 2.0\})$ は N の効果による推定精度の変化を、 $q (\in \{0, 0.1, \dots, 2.0\})$ は N_F の効果による推定精度の変化を確かめるために使用する。

台詞と辞書を用いた BPM 推定手法の予測式を式 (5.3) に示す。目的変数 y は BPM の値である。 c は定数項である。

$$y = a_r x_r + a_b x_b + a_{dp} x_{dp} + a_a x_a + a_s x_s + c \quad (5.3)$$

第6章

評価実験

5章で提案した台詞の重回帰分析によるBPM推定手法の推定精度を検証する。5.1節で作成した特徴語辞書の有効性を確かめるために、2種類の辞書セットを用いて実験を行う。

辞書セット1を図6.1に、辞書セット2を図6.2に示す。日本語シソーラス連想類語辞典[6]で「喜び」「喜ぶ」「嬉しい」「楽しみ」「楽しむ」「楽しい」の6語で検索して出力された語を「喜び/楽しみ」の辞書、「怒り」「怒る」「腹立たしい」の3語で検索して出力された語を「怒り」の辞書、「悲しみ」「悲しむ」「悲しい」の3語で検索して出力された語を「悲しみ」の辞書に登録する。「喜び」と「楽しみ」はどちらも快さを表す類似した感情であるため、1つの辞書にまとめた。「喜び/楽しみ」「怒り」「悲しみ」の感情辞書は、2種類の辞書セットで同じものを用いる。同様に「走る」「走り」の2語で検索して出力された語を登録した辞書と、「戦う」「戦い」の2語で検索して出力された語を登録した辞書を加えたものを辞書セット1とする。一方、シソーラスを基に手作業で作成した辞書の代わりに、5.1節の提案手法によって自動的に作成した「走っている場面」「戦っている場面」の特徴語辞書を加えたものを辞書セット2とする。

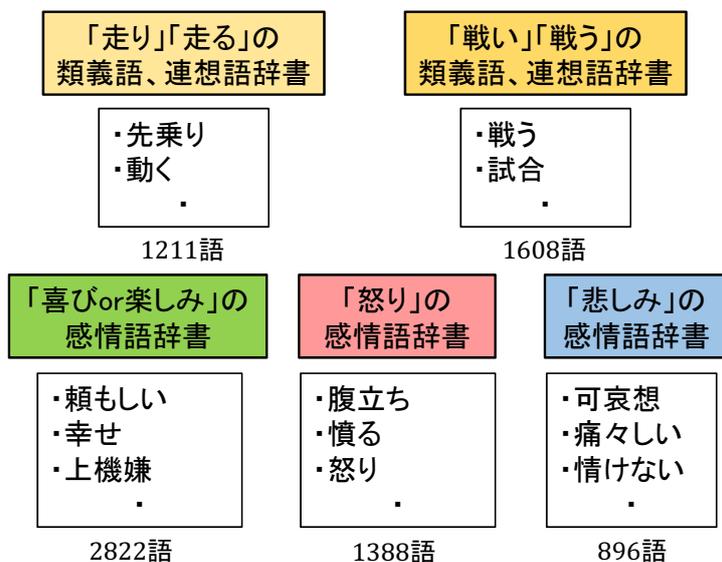


図6.1 辞書セット1（シソーラスを基に手製した辞書）

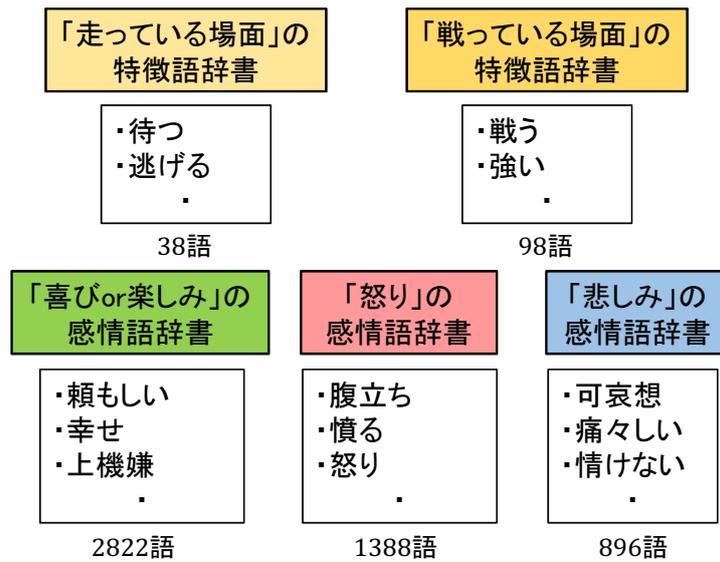


図 6.2 辞書セット 2 (SF-ISF による特徴語辞書を含む)

初めに、2種類の辞書セット毎、正解 BPM と推定 BPM との平均二乗誤差 (MSE) をパラメータ p, q を変化させて比較する。辞書セット 1 における MSE を図 6.3 に、辞書セット 2 における MSE を図 6.4 に示す。サンプルデータは、バトルアニメ 3 作品から 50 場面、ギャグアニメ 2 作品から 40 場面、日常アニメ 1 作品から 10 場面の計 100 個を用いる。MSE が最も小さくなる p, q の組み合わせは、辞書セット 1 では $p = 0.0, q = 1.2$ のとき、 $MSE = 866.3$ であった。一方、辞書セット 2 では $p = 0.8, q = 0.3$ のとき、 $MSE = 675.3$ であり、パラメータ p, q が最適化されている場合、SF-ISF による特徴語辞書を用いた辞書セット 2 の方が MSE が小さいことが分かる。

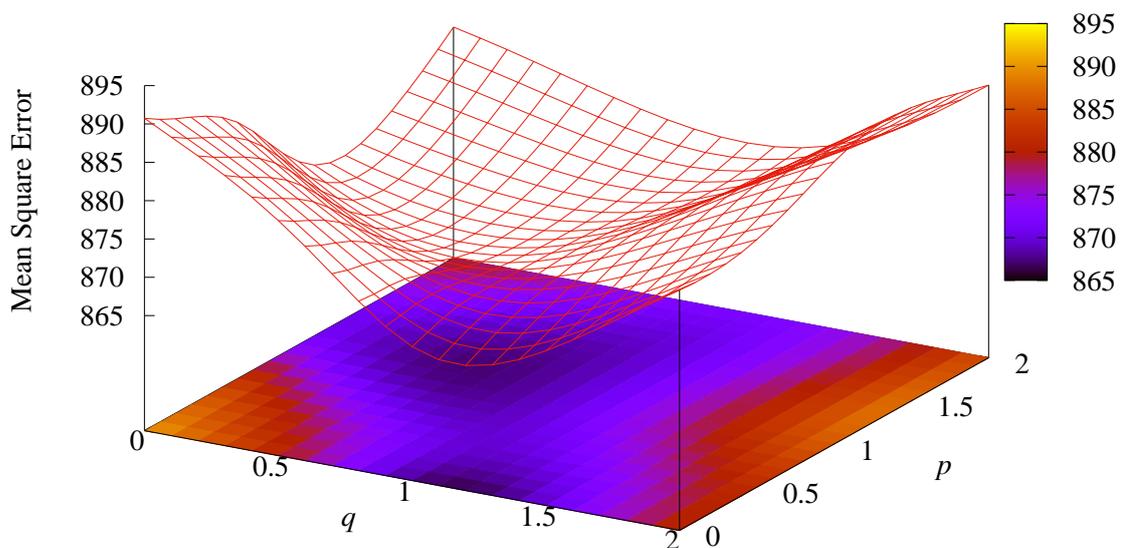


図 6.3 辞書セット 1 (シソーラスを基に手製した辞書) の MSE

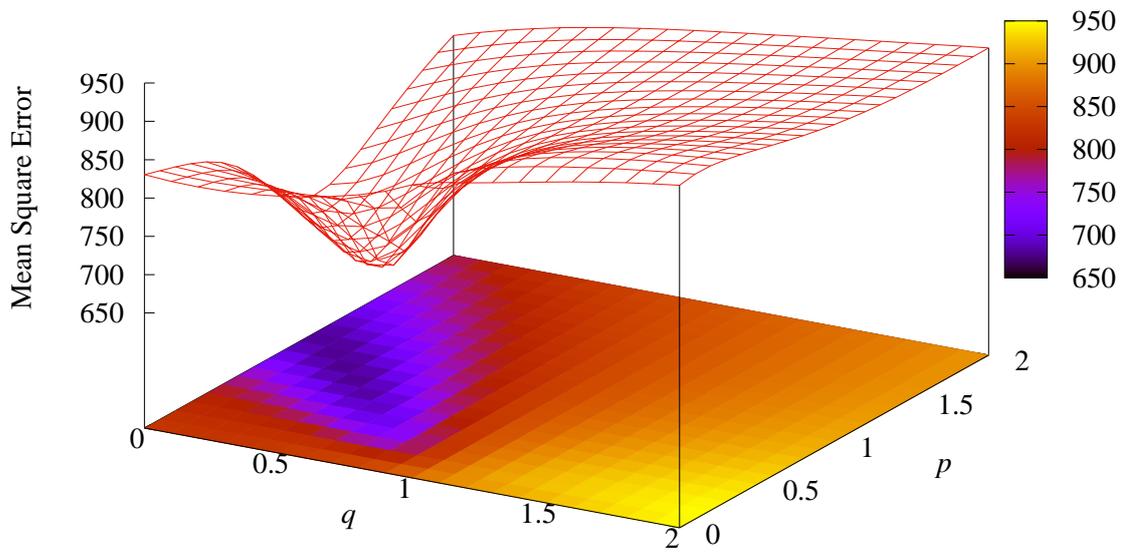


図 6.4 辞書セット 2 (SF-ISF による特徴語辞書を含む) の MSE

次に、辞書セット 1, 辞書セット 2 で各々 MSE が最も小さくなる p, q の組み合わせを用いて、4.3 節と同様に合格率を求める。2 種類の辞書セットにおける許容誤差と合格率の関係を比較するグラフを図 6.5 に示す。許容誤差が 10 前後のとき、辞書セット 2 の方が推定精度が高いことが分かるが、5.1 節で作成した特徴語辞書が、より「走っている場面」「戦っている場面」に出現し易い単語を反映しているからであると考えられる。

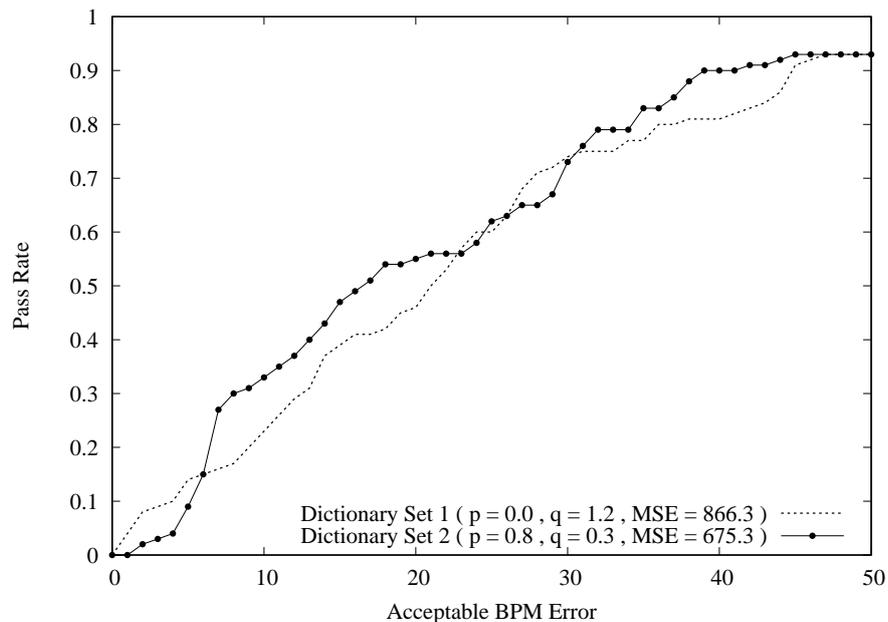


図 6.5 2 種類の辞書セットにおける推定 BPM の合格率

最後に、特徴語辞書の作成における閾値を変化させ、それぞれの閾値における MSE 及び合格率を求める。閾値は 0.00 から 3.00 までの 0.25 刻みの値を使用する。特徴語辞書の作成における閾値を変化させたときの MSE 及び合格率のうち、閾値が 0.00 から 0.75 までの結果を図 6.6 に、閾値が 1.00 から 1.75 までの結果を図 6.7 に、閾値が 2.00 から 3.00 までの結果を図 6.8 に示す。また、閾値と MSE の関係を図 6.9 に示す。図 6.9 では、閾値は 0.00 から 4.00 までの 0.05 刻みの値を使用している。

合格率に関しては、閾値が 1.50 以下の場合、許容誤差が 10 程度のときは合格率が高い傾向にあるが、許容誤差が 40 以上の場合には合格率が低い傾向にあることが分かる。一方、閾値が 1.75 以上の場合、許容誤差が 10 程度のときは合格率が低い傾向にあるが、許容誤差が 40 以上の場合には合格率が高い傾向にあることが分かる。

MSE に関しては、パラメータが最適化されている場合、閾値が 0.95 のときに MSE が最小となり、 $MSE = 587.7$ であった。一方、MSE が最大となったのは閾値が 3.55 以上のときであり、 $MSE = 862.8$ であったが、それでも辞書セット 1 の MSE より小さい。この結果から、閾値によって MSE が大きく変化していることが分かるので、特徴語辞書の作成においては適切な閾値を設定する必要があると考えられる。また、閾値が 0.75 から 0.95 までの間は MSE が減少しているが、これは「走っている場面」や「戦っている場面」に無関係な単語が特徴語辞書に登録されなくなっているからであると考えられる。実際に特徴語辞書を見てみると、SF-ISF の値が 0.75 から 0.95 までの間に「俺」「私」「と」「だ」「は」「を」などの単語があり、閾値を 0.95 まで上げるとこれらの単語が特徴語辞書に登録されなくなる。さらに、閾値が 3.00 を超えると MSE が大きくなったが、これは特徴語辞書に登録される単語数が減少し、台詞中の特徴語が抽出されにくくなったからであると考えられる。実際に特徴語辞書を見てみると、閾値 3.00 の場合の登録語は「走」の辞書で 6 件、「戦」の辞書で 14 件であったのに対し、閾値 3.05 の場合の登録語は「走」の辞書で 2 件、「戦」の辞書で 4 件であり、単語数が大幅に減少している。

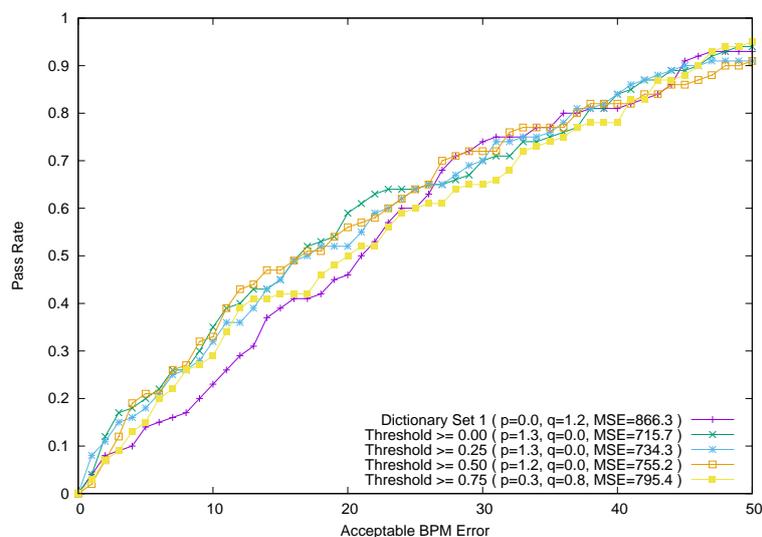


図 6.6 MSE 及び合格率 (閾値 0.00~0.75)

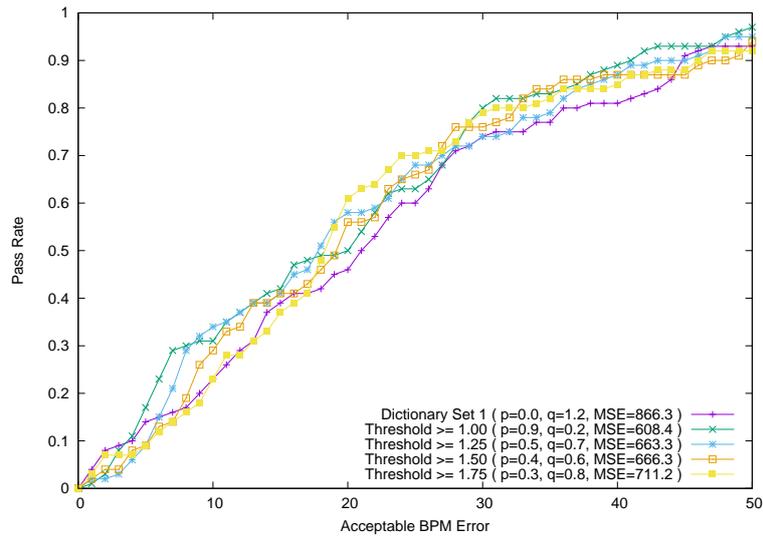


図 6.7 MSE 及び合格率 (閾値 1.00~1.75)

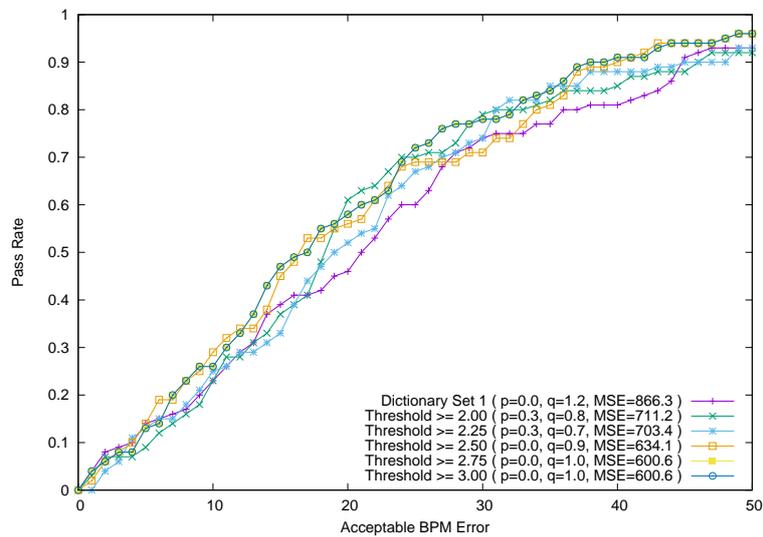


図 6.8 MSE 及び合格率 (閾値 2.00~3.00)

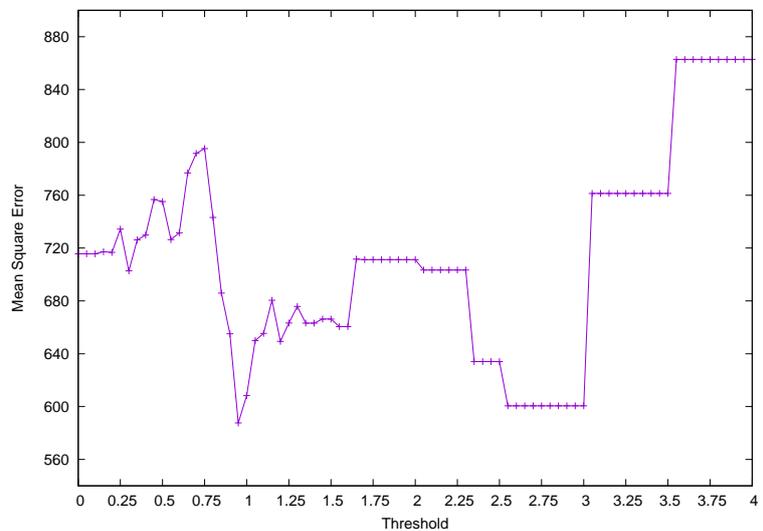


図 6.9 閾値と MSE の関係

第7章

まとめと今後の課題

本研究では、劇伴の自動作曲システムについて、個人制作アニメの場合と会社制作アニメの場合に分けて提案し、劇伴自動作曲の第一歩として劇伴の BPM を推定する手法を検討した。さらに、アニメ映像をシステムへの入力として用いる個人制作アニメを主な対象とし、台詞と辞書を用いた重回帰分析によって BPM を推定する手法を提案した。

BPM に影響を与える要因の検討として、「走っているか否か」と「戦っているか否か」という要因と BPM に一定の相関があることを確認した。さらに、これらの要因を予測式の説明変数に用いて推定 BPM の合格率を求めたところ、あまり高い値にはならなかったため、これら 2 つの要因だけでは不十分であり、他の要因も考慮すべきであることが分かった。

特徴語辞書の作成には、TF-IDF 法を参考にして、特徴語としての重要度を表す SF-ISF (Scene Frequency - Inverse Scene Frequency) を用いた。得られた特徴語辞書から、「待つ」や「逃げる」、「てめえ」「上手い」などといった「走っている場面」や「戦っている場面」を表す特徴語が抽出できていることを確認した。

動作要因と感情を用いた予測式による BPM 推定では、正解 BPM と推定 BPM との平均二乗誤差 (MSE) としては、パラメータ p, q が最適化されている場合、日本語シソーラスを基に手製した辞書セット 1 では $MSE = 866.3$ 、SF-ISF (閾値は「1 より大」) を用いた特徴語辞書を含む辞書セット 2 では $MSE = 675.3$ であり、SF-ISF を用いた特徴語辞書を用いた方が MSE が小さな値となった。推定 BPM の合格率については、辞書セット 2 を用いた方が全体的に高く、おおよそ BPM が推定できていると言える許容誤差 10 程度の場合の合格率も同様に辞書セット 2 を用いた方が高くなった。

閾値に依る MSE の違いを確かめる実験では、閾値が 0.95 のときに MSE が最小となることが分かり、 $MSE = 587.7$ が得られた。一方、MSE が最大となったのは閾値が 3.55 以上のときであり、 $MSE = 862.8$ であった。これより、閾値に依る MSE の差は大きく、特徴語辞書の作成においては適切な閾値を設定する必要があると分かった。

今後の課題としては、まず、評価実験に用いたサンプルデータの数が少なかったため、サンプルデータの数を増やし、個々の作品に依存しない辞書や推定精度を詳しく検証する必要がある。「戦っている場面」の特徴語辞書の上位 10 件の中には「魔」や「秘技」といった単語が含まれており、これらの単語はそれぞれ特定の 1 作品にしか出現していなかった。従って、特定

の作品にしか出現しない語が特徴語辞書に登録されないようにするために、サンプルデータを増やすことが必要である。また、SF-ISF の計算式の検討が不十分であるため、数種類用意して特徴語辞書を作成し、得られた辞書や推定精度を比較する。計算式の例として、パラメータ s, t を用いて $\{SF_p(w)\}^s$ と $\{SF_n(w)\}^t$ とし、 s, t を変化させることを考えている。最後に、アニメ映像から得られる情報には、台詞などのテキスト情報の他に、輝度や輝度変化などの画像情報や、登場人物の話すスピードや抑揚などの音声情報などもあるので、これらの活用方法も検討する必要があると考えられる。

謝辞

本研究に際して、様々なご指導を頂きました服部峻助教に厚く御礼申し上げます。また、日常の議論を通じて多くの知識や示唆を頂いた服部研究室の皆様へ感謝致します。

参考文献

- [1] 深山 覚, 中妻 啓, 米林 裕一郎, 酒向 慎司, 西本 卓也, 小野 順貴, 嵯峨山 茂樹, “Orpheus: 歌詞の韻律に基づいた自動作曲システム,” 情報処理学会 研究報告「音楽情報科学」, Vol.78, No.76, pp.179–184 (2008).
- [2] 藤原 和弘, 高橋 智一, 鈴木 冒人, 青柳 誠司, “遷移確率を用いた自動作曲,” 情報処理学会 研究報告「音楽情報科学」, Vol.2014-MUS-104, No.16, pp.1–6 (2014).
- [3] 古森 優, “プロが教えるアニソンの作り方,” リットーミュージック (2015).
- [4] アニメ劇伴の作り方 その2【劇伴まとめ第四回】, <http://kohrogi.com/?p=9407> (2017).
- [5] アニメ劇伴で使う音楽の種類【劇伴まとめ第一回】, <http://kohrogi.com/?p=9320> (2017).
- [6] 日本語シソーラス連想類語辞典, <http://renso-ruigo.com/> (2017).
- [7] lucene-gosen, <https://code.google.com/archive/p/lucene-gosen/> (2013).