DEIM2022 F24-2(day2 p33)

セリフの感情推定に基づくキャラクタの表情変化の自動化

吉田　裕太†　　服部　　峻††

†,†† 室蘭工業大学 ウェブ知能時空間研究室　〒 050–8585 北海道室蘭市水元町 27–1
E-mail: †21043072@mmm.muroran-it.ac.jp, ††hattori@csse.muroran-it.ac.jp

あらまし　現代のエンターテインメントとして，アニメやゲームなど様々なコンテンツが存在し，これらには多数の魅力的なキャラクタが存在する．キャラクタにおける魅力の観点は様々であり，キャラクタの仕草や表情，見た目，声など数多く存在し，それら全てを一つ一つ手作業で設定していっては，制作側に多大な負担がかかることは容易に想像ができる．この問題に対して著者らは，キャラクタを彩る要素の 1 つとしてキャラクタの表情に着目し，そのキャラクタに与えられるセリフに応じて，そのキャラクタの感情を分析し，キャラクタの表情を適応的に変化させることで，キャラクタのモーション制作においての手間を 1 つ取り除くことはできないかと考えた．本稿では，セリフに含まれる単語から，セリフがどの感情カテゴリに分類されるか，また表れる感情がどの程度の大きさなのかを推定し，数パターンのキャラクタの表情から推定された感情カテゴリと大きさに合うものを選択することで，キャラクタのセリフと表情に違和感が無いように，キャラクタの表情変化を自動的に制御する手法を提案する．

キーワード　感情推定，形態素解析，表情制御，振る舞い制御，ナイーブベイズ分類，ポジネガ判定

1　まえがき

現代のエンターテインメントとして，アニメやゲームなど様々なコンテンツが存在する．それらのコンテンツに共通して登場人物である多種多様なキャラクタが存在する．キャラクタの良さ，人気度などはとても重要であり，それらコンテンツの魅力に大きな影響を与えている傾向があると考えられる．キャラクタにおける魅力の観点は様々であり，キャラクタの仕草や表情，見た目，声など数多く存在し，仕草 1 つにしても手を挙げるや，振り向きなど，それら全てを一つ一つ手作業で設定していっては，制作側に多大な負担がかかることは容易に想像ができる．

この問題に対して著者らは，キャラクタを彩る要素の 1 つとしてキャラクタの表情に着目し，そのキャラクタに与えられるセリフに応じて，キャラクタの表情を自動的に変化させることはできないかと考えた．表情を形成するにあたり，重要な要素はセリフを発しているキャラクタの感情であると考えられる [1]．また，セリフにはそのテキストの内容や，あてられる声などから，そのキャラクタの感情を分析することができると考えられる．セリフに含まれる感情を分析し，キャラクタの表情を適応的に変化させることで，キャラクタ制作においての手間を 1 つ取り除くことができる．

また，最近流行り始めている VTuber というコンテンツに対して，モーションキャプチャのような技術を使用せずとも，後述する Pramook Khungurn [2,3] の「Talking Head Anime from a Single Image」と，本研究のテキスト等のデータから感情を推定し，表情の変化を自動的に制御することで表情変化を自動で行う技術を開発できれば，キャラクタを演じる役者が存在せずとも，キャラクタが自然に動くといったシステムも構築できるのではと考えられる．

本稿では，セリフに含まれる単語から，セリフがどの感情カテゴリに分類されるか，また表れる感情がどの程度の大きさなのかを推定し，数パターンのキャラクタの表情から推定された感情カテゴリと大きさに合うものを選択することで，キャラクタのセリフと表情に違和感が無いように，キャラクタの表情変化を自動的に制御する手法を提案する．

2　関連研究

本研究での実験環境として Pramook Khungurn [2, 3] の「Talking Head Anime from a Single Image」を使用している．これは，膨大な 3D モデルのデータから，顔のパーツの動きを学習しており，単一のキャラクタの画像（256 × 256）を入力として与えた時に，目や口，眉毛などといった顔のパーツを別の形へと変更することで，様々な表情を作成できるという技術（図 1）である．この関連研究は「コストを少なく簡単に VTuber というものになれるようにする」ことを目的としており，正面を向いたキャラクタの画像を入力として与えるだけで，2D モデルのようなキャラクタの動きを表現することができるため，本研究の実験に使用した．

テキストのカテゴリ分類において，Rennie ら [4] の補集合を用いた Complement Naive Bayes（CNB）についての研究や，古宮ら [5] の NB の性質と CNB の性質を合わせた Negation Naive Bayes（NNB）など，テキストに適したカテゴリを推定するために Bayes 手法 [6] はよく用いられる．

本稿では，複数の感情カテゴリに分類する多項分類であり，データセットが不均衡なため Rennie ら [4] の CNB を用いる．ここで，データセットを TwitterAPI [7] から取得したツイートから作成することで，会話口調のテキストや現代語に近いテ

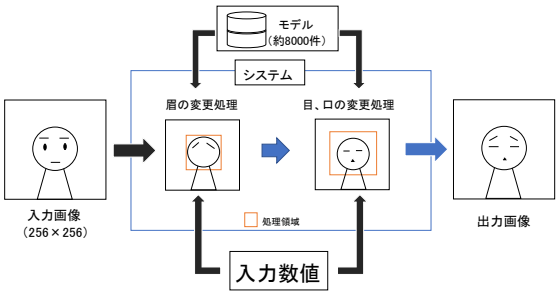

図 1: 関連研究の概観

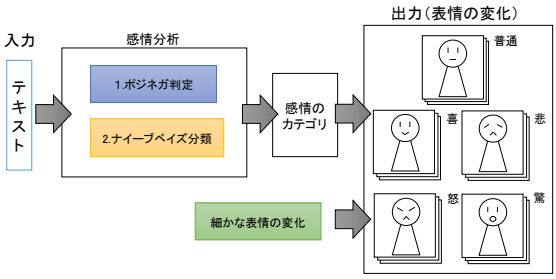

図 2: 提案手法の概観

キストに対しても分類が可能であると考える．データセットの作成などの詳細は，3.3.2 項にて述べる．

3 提 案 手 法

本章では，セリフに含まれる単語から，セリフがどの感情カテゴリに分類されるか，また表れる感情がどの程度の大きさなのかを推定し，数パターンのキャラクタの表情から推定された感情カテゴリと大きさに合うものを選択することで，キャラクタのセリフと表情に違和感が無いように，キャラクタの表情変化を自動的に制御する手法を提案する．

提案手法の概観を図 2 に示す．セリフなど，与えられたテキストを入力として，適した表情を形成するために，まず，テキストを感情分析してテキストに含まれる感情カテゴリを抽出する．次に，テキストに含まれる感情に基づいて，キャラクタの表情を変化させることで，自然な表情の変化を再現する．また，図 2 に示された細かな表情の変化に関しては，入力とは関係無く，変化として存在しないと違和感になるであろうと定義されたものである．詳細は，次の 3.1 節にて述べる．

3.1 細かな表情の動きの制御

表情の変化には，笑っているや泣いているといった大きな変化以外にも，瞬きや発する言葉によって変化する口の動きなどが挙げられる．それらが全く動かないとすると違和感に繋がり，表情の変化を付けたとしても適切な評価に繋がらないと考えられる．そこで，瞬きと口の動きを違和感の無いであろう動きへとあらかじめ設定しておく．

瞬きに関して，人間の瞬きは 1 分間に 10 回から 30 回程度とされている．また，連続で瞬きする可能性も考えられるため，瞬きするタイミングは 1 秒から 6 秒をランダム（一様分布乱数）で選択し瞬きする仕様にしている．

口の動きに関しては，聞き取りやすい話の速さが 1 分間に 300 文字とされているため，1 秒間に 5 文字程度と考え，口の開け閉めを行うものとする．また，テキストを与えた時の口の動かし方がどのようなものが違和感が少ないかを考えると，口の動かし方のパターンとしては，以下の 4 パターンが挙げられると考えられる．どれが違和感の少ないものとして，適切であるかをアンケート形式で被験者を用意し，検証を行った．アンケートの詳細については，4 章の評価実験にて述べる．

- 口を開けたままにさせる
- 口をパクパクと開け閉めをさせる
- 口は開けたままで，テキストの 1 文字ごとの母音に沿って形を変化
- 口を開け閉めしながら，テキストの 1 文字ごとの母音に沿って形を変化

3.2 表 情 制 御

テキストから感情分析して抽出された感情に基づいて表情を制御する手法として，以下の 2 種類の比較手法 1・2，及び，2 種類の提案手法 3・4 を定義し，4 章において比較実験を行って，提案手法の有効性を検証する．

（1） 数個の表情のパターンからランダム
（表情のパターン計 11 個）

（2） 無表情で固定

（3） ポジネガ判定により感情推定した結果を用いたもの
（表情のパターン計 9 個）

（4） ナイーブベイズ分類器により感情推定した結果を
用いたもの（表情のパターン計 9 個）

手法 1 は，作成した複数の表情パターンからランダムに選択する．表情のパターンは 11 個である．

手法 2 は，常に無表情で固定するものである．これは，人間が普段会話する時において表情がそこまで変化が起こらない場合があるとも考えられるため，表情に変化が起きなくても違和感が生じないのではと考えたためである．一方で，違和感は生じないが，それが魅力的であるとは考えられないため，生き生きしているかという評価尺度では低い結果となると予想される．

手法 3 と 4 は，感情分析を用いて表情を変化させる手法である．手法 3 は，ポジネガ判定により感情を推定し，それに伴って表情を変化させるものである．ポジネガ判定により，ポジティブとネガティブに分類することで，大まかな感情の分類を行う．分類後，あらかじめポジティブとネガティブに分類しておいた表情のパターンの中から，ランダムで 1 つ選択し出力する．例として，正解が「悲しい」というネガティブに分類する

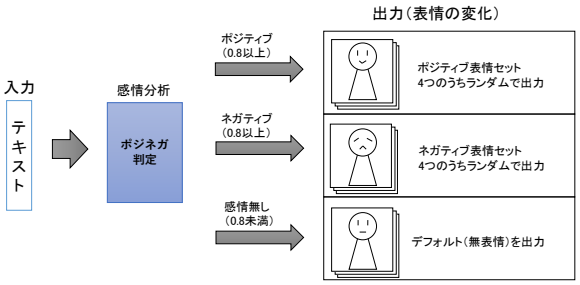

図 3: ポジネガ判定による表情の変化の概要

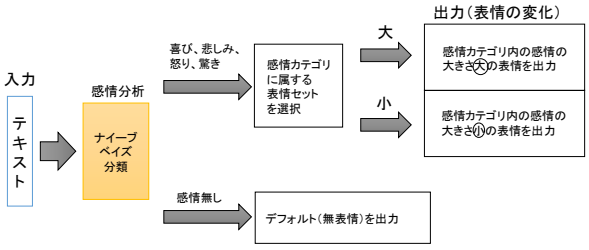

図 4: ナイーブベイズ分類による表情の変化の概要

ものでも，同じネガティブに分類される「怒り」の表情でも違和感が生じない場合があると考えられ，ポジティブとネガティブという大きな枠組みであっても，表情の変化に違和感が生じない可能性がある．作成した表情セットは合計 9 個であり，ポジティブが 4 つ，ネガティブが 4 つとなっている．また，感情無しと推定された時のデフォルトの表情も 1 つ含む．手法 3 の概要を図 3 に示す．

手法 4 は，「喜び」，「悲しみ」，「怒り」，「驚き」の 4 つの感情においてナイーブベイズ分類を用い，感情カテゴリの推定と，ナイーブベイズによる確率から感情の大小を推定することで，表情の変化を制御するという手法である．ポジネガ判定よりも細かく感情カテゴリを分割し，より正確な感情の推定を行うことで違和感が生じる可能性を減らすことができると考えられる．作成した表情セットは合計 9 個である．それぞれの感情カテゴリに 2 つずつ存在し，そのうちの感情の大小で 1 つずつとなる．また，感情無しと推定された時のデフォルトの表情も 1 つ含む．手法 4 の概要を図 4 に示す．

それぞれに属する表情セットは，同じ表情を含んでいるものも存在する．手法 3 のポジネガ判定のネガティブに属する表情セット 4 つは，手法 4 のナイーブベイズ分類器における「悲しみ」と「怒り」の合計 4 つと同じである．また，ポジネガ判定のポジティブの表情セットの中には 4 つ存在するが，その中の 2 つは，ナイーブベイズ分類器の「喜び」の 2 つと同じ表情であり，残り 2 つはポジティブ用に新たに作成された表情である．そして，手法 1 の表情セット 11 個は，ナイーブベイズ分類器に用いた 8 個と，ポジネガ判定のポジティブ用に作成された 2 つを合わせた 10 個に，デフォルトの無表情 1 つを加えたものである．手法 3 と手法 4 における感情分析の詳細に関しては 3.3 節で述べる．

3.3 感情分析

感情推定の手法として，テキストのポジネガ判定を用いた感情推定方法と，ナイーブベイズを用いて感情の分類器を作成，使用する感情推定方法の 2 つを試行する．

3.3.1 ポジネガ判定による感情推定手法

感情分析において主流であるポジネガ判定を用いる．用いる手法は，python のライブラリに存在する asari を使用する．辞書ベースではなく，scikit-learn を利用し文章を tf-idf でベクトルへと変換し，SVM を用い分類問題としてポジティブであるか，ネガティブであるかを分類するというものである．辞書ベースと違い未知語にも強く，自然言語処理において精度が良いとされている BERT と比較しても遜色ない精度を発揮し，BERT よりも処理がわずかに早いという点から，入力としてテキストが与えられた時に，表情の変更までのラグを減らすことができ，入力から出力までのラグが違和感に繋がるということを避けられるため，本研究に適していると判断した．

ポジネガ判定によるポジティブとネガティブへの分類の他に，感情無しへの判定を作成するためポジティブとネガティブへの各々の判定確率を用い，ポジティブ・ネガティブと感情無しを分ける閾値を設定する．ここで，使用するデータは TwitterAPI から取得したテキストを第 1 著者が確認し，感情が含まれると判断したツイート 100 件を使用する．取得した 100 件を asari にかけ，ポジティブとネガティブに推定される確率を求める．その高いスコアの方を取得し，全ての平均を算出する．算出した結果，平均は 0.80 となったため，ポジティブ・ネガティブと感情無しを分ける閾値は 0.80 とする．

3.3.2 ナイーブベイズ分類器による感情推定手法

本稿におけるナイーブベイズ分類器は，人間の手によって作られた教師データを用い，単語の出現頻度を学習する．その後，入力テキストに含まれる単語を解析し，学習した単語の頻度からカテゴリに分類される推定確率を求め，カテゴリに分類するというものである．本稿でのナイーブベイズ分類器の作成の大まかな流れは以下の通りである．

（1） TwitterAPI から指定したキーワードを含むツイートを取得

（2） 取得したツイートに前処理を行い，データセットを作成

（3） データセットからモデルを作成

（4） 感情有りと感情無しを分ける閾値の設定

本稿では，学習データに TwitterAPI から取得したツイートを著者がラベル付けを行って用いる．TwitterAPI から取得するツイートには，指定されたキーワードを含むツイートを取得する「キーワード検索」を用いる．キーワードは，感情表現の言葉である「喜び」，「悲しみ」，「怒り」，「驚き」の 4 つをベース

表 1: ツイート取得に用いたキーワード

喜び	嬉しい，楽しい，幸せ，気持ちいい，スッキリ，笑い， 満足，爽快，感動，感心，和む，癒される，落ち着く， ワクワク，興奮する，高ぶる，懐かしい，愉快，快楽， 楽しみ
悲しみ	憂鬱，失望，虚しい，情けない，喪失感，惨め，つらい， へこむ，がっかり，屈辱，切ない，泣ける，苦労，萎える， かわいそう，寂しい，悲しい，孤独，困る，哀れ
怒り	不愉快，不機嫌，反感，批判，怒鳴る，呆れる， イライラ，険しい，腹立，激怒，キレる，八つ当たり， 鬱憤，ブチギレ，激おこ，うざい，マジギレ，イラつく， プンプン，むかつく
驚き	慌てる，予想外，焦る，息苦しい，戦慄，鳥肌が立つ， 不思議，怪奇，びっくり，驚愕，仰天，驚，感嘆， ショック，衝撃，愕然，一驚，意表，驚嘆，不可思議

に，それらの類義語をキーワードとして指定する．また，キーワードで取得したツイートの感情ラベルは，ベースとなった 4 つの感情カテゴリが付く．キーワードは表 1 に示す．キーワードの個数は 20 個とし，キーワード 1 個につき 1000 件のツイートを取得する．取得したツイートの件数は，完全一致したツイートを 1 件として，「喜び」15221 件，「悲しみ」14927 件，「怒り」9249 件，「驚き」10762 件の計 50159 件である．

次に，取得したツイートに対して，前処理を行う．前処理は以下の 3 点である．

- 感情を含むツイートと含まないツイートに分類
- 指定した品詞に該当する単語を取得
- 否定語の判定

キーワードを用い取得したツイートには，キーワードの単語のみのテキストや，感情が含まれていないと思われるツイートが含まれる可能性があるため，取得したツイートそのままでは，問題があると考えられる．そのため，ツイートに感情が含まれるかどうかを判別するために，asari のポジネガ判定に一度かけスコアを出すことで，スコアの値がある閾値を満たさないものは除くという処理を行う．3. 3. 1 項と同様に，閾値は 0.8 とした．ポジティブまたはネガティブのスコアが 0.8 以上の場合は感情有りのツイートとして取得し，0.8 を下回った場合は感情無しのツイートとして除外する．この前処理によって，取得したツイートは，「喜び」11604 件，「悲しみ」7279 件，「怒り」4192 件，「驚き」5978 件の計 29053 件となり，選抜ツイートと呼ぶ．

取得した選抜ツイートのテキストから特徴量を抽出するために，単語の出現回数をカウントし特徴量とする scikit-learn の Countvectorizer を用いる．ここで，カウントする単語の品詞を指定し，該当するものをカウントする．これは，助詞や接続詞のような，どの文にも含まれている可能性のある単語を除き，ノイズを減らすために行う．取得する品詞を決定するために，それぞれの品詞で単語を取得した時の精度を比較する．精度比較に用いたテストデータは，TwitterAPI から学習データ作成時と同じ手法を用い作成した 10000 件のデータを用いる．また，本稿で比較する品詞は「名詞」,「形容詞」,「動詞」,「副詞」の 4 つである．副詞を含む理由としては，取得したツイートの感情有りとされた選抜ツイートの中に副詞のみで構成されたテキストが含まれており，副詞にも感情が含まれる単語が存在するのではと考えたためである．結果としては，表 2 となった．「名詞」,「形容詞」,「動詞」,「副詞」の 4 つ全てを含んでいる場合の accuracy が一番高いため，品詞は「名詞」,「形容詞」,「動詞」,「副詞」を用いる．

表 2: 取得した品詞ごとのナイーブベイズ分類の精度

品詞	名	動	形	副
accuracy	0.58	0.35	0.40	0.28

品詞	名・動	名・形	名・副	動・形
accuracy	0.64	0.67	0.60	0.48

品詞	動・副	形・副	名・動・形	全て
accuracy	0.38	0.42	0.72	0.74

表 3: 否定語判定のための語句

助動詞	ない，なかろ，なかっ，なく，なけれ，ぬ，ず，ん，ね
助詞＋助動詞	じゃ　＋　ない
接頭辞	非，不，無，未，反，異

次に否定語の判定について述べる．否定語の判定には，パターンマッチを用いる．動詞や名詞などの後に続いて，否定形の助動詞などが存在する場合，それらをまとめて一つの単語としてカウントする．これにより，ある程度の否定語に対応することができると考えられる．パターンマッチで検出する単語は表 3 に示す．

前処理が終わり，次にモデルを作成する．本稿では，分類するカテゴリが複数存在し，作成したデータセットが感情カテゴリごとに数が異なるため，ナイーブベイズ分類のモデル作成時に利用する scikit-learn の complementNB を用いる．これは，不均衡なデータセットに適したアルゴリズムであり，本稿のデータセットに適していると考えられる．

以上がモデル作成までの流れである．ナイーブベイズ分類器を用いた日本語のテキストに対しての感情推定方法は，テキストに含まれる単語 1 つずつに感情カテゴリごとの感情推定確率を求め，テキスト全体の平均を求める．平均が一番高い感情カテゴリをそのテキストが表す感情としている．本稿で作成したモデルの精度は表 4 に示す．精度としては，0.51 と決して高くはないため，感情分析の精度向上は今後必要であると考えられる．

ナイーブベイズ分類においても，感情有りと感情無しを分けるための閾値を設定する．3. 3. 1 項のポジネガ判定による感情推定手法において用いた感情有りツイート 100 件をナイーブベイズ分類器にかけ，感情カテゴリを推定した時の感情推定確率を用いる．データ 100 件における推定した感情カテゴリの推定確率の分布は図 5 のようになる．図 5 の平均が 0.28 であり，これをナイーブベイズ分類器における感情有りと感情無しに分け

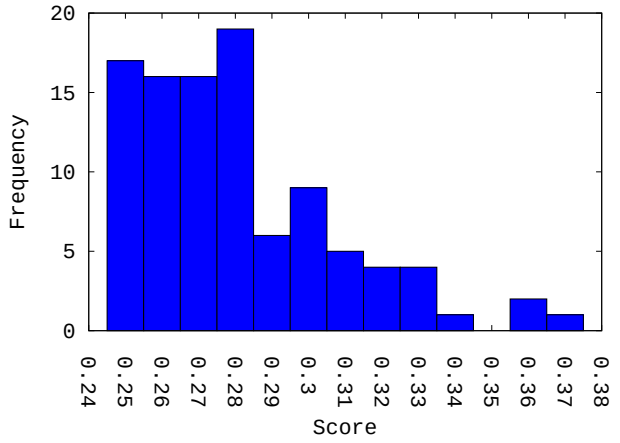

図 5: $N = 100$ の時の感情推定確率の分布

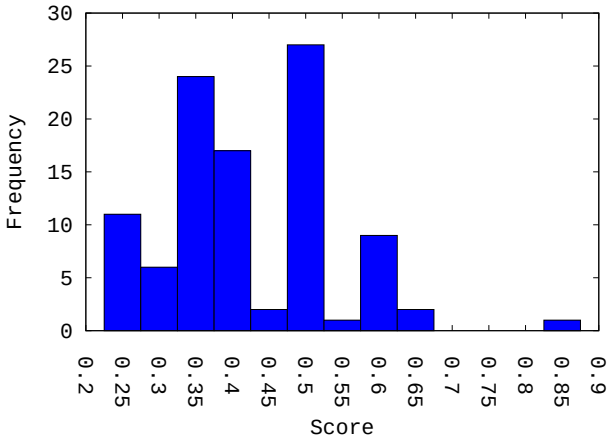

図 6: $N = 100$ の時の感情推定確率の最大値の分布

表 4: ナイーブベイズ分類器の精度

	Precision	Recall	F1
喜び	0.59	0.67	0.63
悲しみ	0.52	0.43	0.47
怒り	0.35	0.47	0.40
驚き	0.56	0.38	0.45
全体の精度			0.51

る閾値とする．

また，ナイーブベイズ分類においては，感情の有無の他に感情の大小を決める閾値も設定する必要がある．同じく，データ 100 件を用いそれぞれの感情推定確率を求め，感情カテゴリの推定を行う．感情の有無の判定とは違い，感情の大小を決める際に参照するのは，感情カテゴリを推定し，テキストに含まれる単語ごとの推定したカテゴリの感情推定確率を取得し，それらの最大値を用いる点である．これは推定した感情において，テキストに含まれる単語の中で最も強く，その感情を表現している単語がどれくらいの大きさなのかによって，感情の大小を測るためである．データ 100 件における推定した感情カテゴリの推定確率の最大値の分布は図 6 に示す．図 6 の平均が 0.44 であり，これをナイーブベイズ分類器における感情の大小を分ける閾値とする．閾値を用いたナイーブベイズ分類器による推定の例を図 7 に示す．

4 評価実験

本章では，セリフに含まれる単語から，セリフがどの感情カテゴリに分類されるか，また表れる感情がどの程度の大きさなのかを推定し，数パターンのキャラクタの表情から推定された感情カテゴリと大きさに合うものを選択することで，キャラクタのセリフと表情に違和感が無いように，キャラクタの表情変化を自動的に制御するという提案手法の有効性を検証する．検証する項目は以下の 2 つである．

- テキスト入力時の口の動き方について
- 感情推定を用いた時の表情変化について

この 2 つについて，被験者に対しアンケートを取るという形で

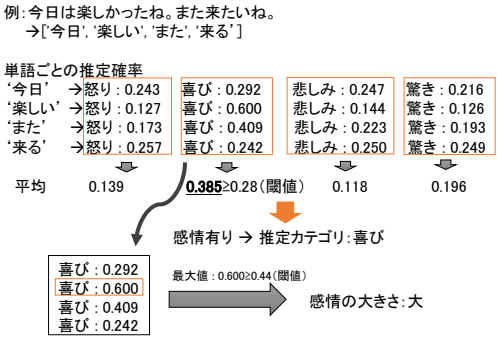

図 7: 閾値を用いたナイーブベイズ分類の推定

実験を行う．実験の環境として，Pramook Khungurn [2,3] の「Talking Head Anime from a Single Image」を使用し，1 枚の画像から顔のパーツごとに動かすことで，キャラクタの表情の変化を再現する．

4.1 口の動き方に関する検証

3.1 節にて述べた，口の動きの 4 パターンを以下に示す．

（1） 口を開けたままにさせる

（2） 口をパクパクと開け閉めをさせる

（3） 口は開けたままで，テキストの 1 文字ごとの母音に沿って形を変化

（4） 口を開け閉めしながら，テキストの 1 文字ごとの母音に沿って形を変化

実験方法としては，キャラクタに 4 パターンの口の動きをさせたものを動画として撮り，被験者 5 名に 4 つの動画を視聴してもらう．4 つの動画のキャラクタの口の動きを比較してもらい，どれが自然であるかという点において相対的な順位付けでの評価と，自然的に見える度合いから絶対的な 5 段階評価を行ってもらった．順位付けの結果は図 8，5 段階評価の結果は表 5 のようになった．

どちらの結果からも，母音に沿って口を開け閉めする動かし方であるパターン 4 が一番自然であるという結果である．ここ

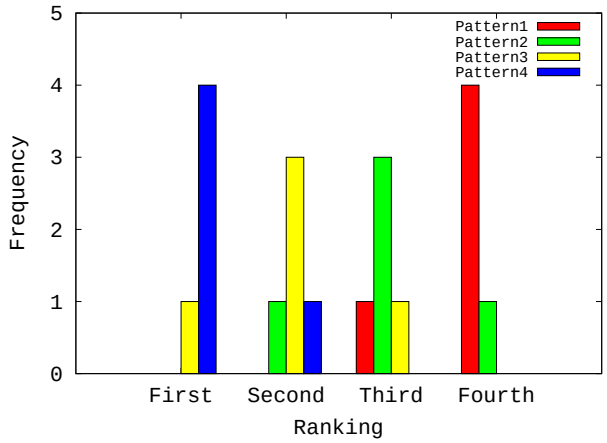

図 8: 口の動き方に関する順位付けの結果（$N = 5$）

表 5: 口の動き方に関する 5 段階評価の結果（$N = 5$）

	1	2	3	4	5	平均
パターン 1	4	1	0	0	0	1.2
パターン 2	1	1	3	0	0	2.4
パターン 3	0	1	3	0	1	3.2
パターン 4	0	0	0	5	0	4.0

表 6: 口の動き方に関する 5 段階評価の平均の有意差検定の p 値（$N = 5$, *：$p < 0.05$）

	比較対象	p 値
口の開閉の有無	パターン 1 とパターン 2	* 0.036
	パターン 3 とパターン 4	0.178
母音による変更の有無	パターン 1 とパターン 3	* 0.013
	パターン 2 とパターン 4	* 0.016

で，有意水準 0.05 とした時の平均の有意差検定 [8] の p 値を求め，結果を表 6 に示す．

結果として，母音に沿って形を変化させた場合はどちらも有意差が見られたが，口の開閉の有無に関しては，パターン 3 とパターン 4 では有意差が見られなかった．これは，母音による口の変化の有無が大きく作用しており，パターン 3 とパターン 4 を比較した時，母音の「い」,「う」の口の形が少し口を小さくするため，口が閉じているように見えたということが考えられる．そのため，口の開閉のみのパターン 1 とパターン 2 では有意差が見られたと考えることができ，口の開閉も自然に見せるために必要であると考えられる．

よって，母音による口の変化と，口の開閉どちらもキャラクタの口の動きとして必要であると言える．この結果から 4.2 節の実験では，キャラクタの口の動きは，母音による口の変化と，口の開閉の 2 つを行うパターン 4 に基づいて作成されている．

4.2 感情推定を用いた表情変化の検証

評価尺度には，キャラクタの表情の変化がテキストと違和感が無いかの「自然さ」，及び，キャラクタが「生き生きしているか」の 2 点を用いる．実験方法としては，いくつかのテキストが与えられた時のキャラクタの表情の変化を被験者 5 名に見てもらい，実験パターン 5 個分に対し評価尺度 2 点に関して，5 段階で評価してもらう．パターンは以下の 5 つである．

表 7: 実験にて使用したセリフ

	感情カテゴリ	セリフ
テキスト 1	喜び	今と変わらずかわいいってことだよ！
テキスト 2	怒り	邪魔しないでください。
テキスト 3	悲しみ	行かないでよ，また一人ぼっちになっちゃう。

表 8: テキストごとの感情推定結果

	テキスト 1	テキスト 2	テキスト 3
ポジネガ	ポジティブ	ネガティブ	感情無し
ナイーブベイズ	喜び	感情無し	怒り

（1） 数個の表情のパターンからランダム

（2） 無表情で固定

（3） ポジネガ判定により感情推定した結果を用いたもの

（4） ナイーブベイズにより感情推定した結果を用いたもの

（5） 理想の表情の変化を自作したもの（一番正解に近い）

理想の表情の変化は，セリフとしてテキストを用意する際にアニメのセリフを参考にすることで，そのセリフを発したキャラクタの表情の変化と可能な限り似せて第 1 著者が作成した．よって，正解に近いものとする．本稿で用意したセリフは表 7 の 3 つである．セリフと付随している感情カテゴリは，セリフ取得時のキャラクタの表情を見て，第 1 著者が感じた感情である．また，テキストごとの感情推定結果は表 8 のようになった．ポジネガ判定は，テキスト 3 に対して閾値未満であったため感情無しと誤判定になっており，ナイーブベイズはテキスト 2 と 3 に対して別のカテゴリに分類されており，誤判定となってしまった．

実験としては，3 つのテキストをそれぞれ入力として与えた時のキャラクタの表情の変化を動画として撮り，被験者 5 名に 2 つの評価尺度について 5 段階評価を行ってもらった．動画の見せ方としては，テキストを 1 つずつの動画として見せた場合と，3 つのテキストの表情の変化を繋げて 1 つの動画として見せた場合の 2 パターンで実験を行う．目的として，1 つずつ見せた場合では手法ごとの表情の変化がどれくらい自然であるかを主に見るためである．また，3 つのテキストを 1 つの動画として見せる場合では，表情の変化の動きから生き生きしているかを主に見ることができると考えられる．初めに，テキスト 1 つずつを動画として見せた場合の「自然さ」に関する結果を表 9 から表 11 に示す．

5 段階評価の結果として手法ごとの平均を比較すると，手法 3 のポジネガ判定を用いた手法が，どのテキストにおいても高いという結果になった．ポジネガ判定によってポジティブとネガティブに分類することで，その表情パターンの中の正解と近しい表情が出力されたためである．実際に，テキスト 2 において正解の感情カテゴリは「怒り」で，表示された表情は「悲し

表 9: テキスト 1 の「自然さ」についての 5 段階評価の結果（$N = 5$）

	1	2	3	4	5	平均
手法 1	0	1	1	2	1	3.6
手法 2	3	1	0	1	0	1.8
手法 3	0	0	0	4	1	4.2
手法 4	0	2	1	2	0	3.0
正解	0	0	1	3	1	4.0

表 10: テキスト 2 の「自然さ」についての 5 段階評価の結果（$N = 5$）

	1	2	3	4	5	平均
手法 1	3	2	0	0	0	1.4
手法 2	3	1	1	0	0	1.6
手法 3	0	1	1	3	0	3.4
手法 4	1	2	2	0	0	2.2
正解	0	0	1	2	2	4.2

表 11: テキスト 3 の「自然さ」についての 5 段階評価の結果（$N = 5$）

	1	2	3	4	5	平均
手法 1	1	3	0	1	0	2.2
手法 2	2	2	1	0	0	1.8
手法 3	1	2	1	1	0	2.4
手法 4	2	1	1	1	0	2.2
正解	0	0	1	2	2	4.2

み」の表情であったが違和感はあまり感じられなかったようである．このように，セリフに依るところはあるが，ポジティブとネガティブのカテゴリ分類だけでも正しければ，ある程度違和感が無く，自然に見えるのではないかと考えられる．また，手法 4 のナイーブベイズ分類の評価が良くなかったのは，テキスト 2 とテキスト 3 において誤判定してしまい，違和感のある表情を出力してしまったという点と，テキスト 1 において，感情カテゴリの分類は正しくても，感情の大きさの観点で違和感が出てしまったのではないかと考えられる．

次に，テキストごとの結果から，アンケートの 5 段階評価を用いて表情の変化の「自然さ」について統計的に比較する．それぞれの手法の比較に有意差検定を行い比較し，正解とそれぞれの手法の比較に同等性検定 [9] を行い比較する．

有意差検定では，提案手法 2 つの比較と，提案手法それぞれと比較手法の 2 つを比較する．結果は表 12 のようになった．有意差が見られたのは，テキスト 1 における手法 3・手法 4 と，テキスト 2 における手法 3・手法 1，手法 3・手法 2 の 3 点であり，手法 3 に関する箇所で有意差が見られた．また，手法 3 に関する p 値はテキスト 3 を除き，有意差が見られないものの有意水準 0.05 に近い値も見られるため，手法 3 は本稿での手法において最も自然であるという結果になった．

さらに，正解に対するテキストごとの「自然さ」についての 5 段階評価から，許容誤差 $|\epsilon| \leqq 0.5$ で同等性検定を行うと，p 値や 95%同等信頼区間の下側信頼限界（LCL）と上側信頼限界（UCL）は表 13 のようになった．どの手法も p 値が有意水準 0.05 を超えており，正解との同等性は認められなかった．しかしながら，どのテキストにおいても p 値が最も低いものは手法 3 であった．このことから，本稿での手法において最も表情の変化が自然に見える手法は，手法 3 のポジネガ判定という結果になった．

表 12: テキストごとの「自然さ」についての 5 段階評価の平均の有意差検定の p 値（$N = 5$, *：$p < 0.05$, **：$p < 0.01$）

	テキスト 1	テキスト 2	テキスト 3
手法 3 と手法 4	* 0.050	0.060	0.803
手法 3 と手法 1	0.323	** 0.004	0.784
手法 3 と手法 2	0.114	* 0.013	0.374
手法 4 と手法 1	0.402	0.117	1.000
手法 4 と手法 2	0.146	0.305	0.582

表 13: 正解に対するテキストごとの「自然さ」についての 5 段階評価の平均との同等性検定（$N = 5$）

	テキスト 1			テキスト 2			テキスト 3		
	LCL	UCL	p 値	LCL	UCL	p 値	LCL	UCL	p 値
手法 1	-0.74	1.54	0.44	1.95	3.65	1.00	0.84	3.16	0.98
手法 2	0.92	3.48	0.98	1.58	3.62	1.00	1.42	3.38	1.00
手法 3	-0.91	0.513	0.22	-0.22	1.82	0.70	0.61	2.99	0.96
手法 4	-0.03	2.03	0.80	1.02	2.98	0.99	0.68	3.32	0.97

次に，3 つのテキストをそれぞれ与えた時の表情の変化を，1 つの動画にまとめたものを被験者 5 名に見せた時の「生き生きしているか」についての 5 段階評価の結果を表 14 のようになった．また，統計的に見るため手法ごとの比較に有意差検定を行い，正解に対する手法ごとの比較に許容誤差 $|\epsilon| \leqq 0.5$ で同等性検定を行った．その結果は表 15 と表 16 になった．

表 14 から「生き生きしているか」についても，手法の中で比較すると手法 3 が最も高い結果である．また，手法 1 と手法 4 もそこそこ高い結果であり，手法 2 が他と大きく差が開いて低いという結果になった．

有意差については，手法 3 と手法 4 のどちらにおいても，手法 2 とは p 値が有意水準 0.05 を下回っており有意差が見られるが，他とは有意差が見られなかった．この結果から，やはり手法 2 のように表情の変化がないものよりも，表情の変化がある他の手法の方が良い評価であり，「生き生きしているか」については，表情の変化があるかどうかが重要であるという結果であった．

さらに正解と比較した時の手法ごとの同等性については，どの手法においても p 値が有意水準 0.05 を超えており，同等性は確認できなかった．しかしながら，p 値が最も低い手法は手法 3 であり，「生き生きしているか」についても，手法 3 のポジネガ判定を用いたものが最も高い評価であったと確認できた．

よって，「自然さ」と「生き生きしているか」の 2 つの評価尺度で最も評価が高かった手法は，手法 3 のポジネガ判定による感情推定を用いて表情の変化を制御したものであった．

表 14: 3 つのテキストをまとめた時の「生き生きしているか」についての 5 段階評価（$N = 5$）

	1	2	3	4	5	平均
手法 1	1	1	1	2	0	2.8
手法 2	4	1	0	0	0	1.2
手法 3	0	0	3	1	1	3.6
手法 4	0	1	4	0	0	2.8
正解	0	0	0	2	3	4.6

表 15: 3 つのテキストをまとめた時の「生き生きしているか」についての 5 段階評価の平均の有意差検定の p 値（$N = 5$）

	p 値
手法 3 と手法 4	0.123
手法 3 と手法 1	0.295
手法 3 と手法 2	** 0.002
手法 4 と手法 1	1.000
手法 4 と手法 2	*** 0.000

* : $p < 0.050$
** : $p < 0.010$
***: $p < 0.001$

表 16: 正解に対する 3 つのテキストをまとめた時の「生き生きしているか」についての 5 段階評価の平均との同等性検定（$N = 5$）

	LCL	UCL	p 値
手法 1	0.54	3.06	0.95
手法 2	2.81	3.99	1.00
手法 3	0.10	1.90	0.84
手法 4	1.21	2.39	1.00

5 まとめと今後の研究課題

本稿では，キャラクタの表情の変化の自動化を行うために，キャラクタが発するセリフに含まれる単語から感情カテゴリと感情の大きさを，ポジネガ判定とナイーブベイズ分類の 2 つの手法を用いて推定し，数パターン作成したキャラクタの表情の中から，推定した感情カテゴリと感情の大きさに合うものを選択することで，キャラクタの表情の変化とセリフに違和感が無いように表情の変化を制御する手法を提案し，アンケートによる検証を行った．

その結果，2 つの評価尺度の「自然さ」と「生き生きしているか」のどちらにおいてもポジネガ判定を用いて感情推定を行い，ポジティブ・ネガティブそれぞれに該当する表情のパターンを出力するという手法が，他の手法と比較して最も高い評価であった．これは，同じネガティブに属する「怒り」と「悲しみ」において，正解が「怒り」で，推定したものがネガティブの時，その内の「悲しみ」の表情が出力されても違和感に感じない場合があるというように，ポジティブ・ネガティブの推定が正しいと，その中で違った表情が出力されても，違和感に感じにくい場合があるためである．このことから，ポジネガ判定のような，感情分析の 2 値分類でもセリフと表情の変化に，違和感を無くすことができ，ある程度は自然に見えるようになったのではないかと考えられる．

ナイーブベイズ分類に関しては，感情カテゴリが細分化されているため推定が正しければ，ポジネガ判定よりも高い評価を得られ，また，多項分類が可能であれば，感情カテゴリをより細分化することで，表情の出力にバリエーションを持たすことができ，セリフとの対応関係がより増え，自然さが確保できると考えていた．しかしながら，4 つのカテゴリという多項分類においてでも推定精度が悪く，また学習データに無い単語が多い文には感情無しと判定してしまうように，現状では誤判定が目立ってしまうため，上手く作用しなかった．

今後の課題としては，感情分析の推定精度の向上が挙げられるが，かなり困難なため，どれくらいの推定精度を確保できれば，表情の変化とセリフに違和感を無くすことができるのかを検証し，必要な推定精度の指標を求めたいと考えている．

また，本稿ではセリフの前後関係等は考慮しておらず，与えられたセリフを発している時の表情を変化させるだけであった．それでは，1 つ前のセリフでは怒っていたのに，次のセリフでは急に笑い出すといった違和感が感じられるものになってしまう．そのため，今後はいくつかのセリフを与えた時に，前後の文を考慮して感情の変化を推定し，表情の変化を行うことで，より違和感を取り除くことを考えている．

さらに，本稿ではセリフのテキストに対して，感情分析を行い表情の変化の自動化を試みたが，テキスト以外の要素に関しても考慮していきたいと考えている．例として，アニメやゲームにはキャラクタの音声が存在し，その音声の大きさなどから感情の大きさ等を推定し，それに基づいてキャラクタの表情を変更することである．テキスト解析だけでは分析できないことについて，キャラクタの音声のような要素を追加することで，より自然的なキャラクタの表情の変化の自動化を目指して，今後の研究を進めて行く．

文　献

[1] 髙木 幸子, “対人場面における顔表情の役割,” 日本心理学会第 70 回大会 (2006).

[2] Pramook Khungurn, “Talking Head Anime from a Single Image,” (2019).

[3] Pramook Khungurn, “Talking Head Anime from a Single Image 2:More Expressive,” (2021).

[4] Jason D. M. Rennie, Lawrence Shih, Jaime Teevan, David R. Kager, “Tackling the Poor Assumptions of Naive Bayes Text Classifiers,” Proceedings of the ICML2003, pp.616–623 (2003).

[5] 古宮 嘉那子, 伊藤 裕佑, 佐藤 直人, 小谷 善行, “文書分類のための Negation Naive Bayes,” 自然言語処理, Vol.20, No.2, pp.161–182 (2013).

[6] “sklearn を使用して規制対象の業界でナイーブベイズアルゴリズムを使用する理由と方法 — Python + コード,” https://ichi.pro/sklearn-o-shiyoshite-kisei-taisho-no-gyokai-de-nai-bubeizuarugorizumu-o-shiyosuru-riyu-to-hoho-python-ko-do-241721569620687

[7] Twitter API, https://developer.twitter.com/en/docs/twitter-api (2021).

[8] Student (W. S. Gosset), “The Probable Error of a Mean, Biometrika,” Vol.6, No.1, pp.1–25 (1908).

[9] D. J. Schuirmann, “A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability,” Journal of Pharmacokinetics and Biopharmaceutics, Vol.15, No.6, pp.657–680 (1987).