

令和2年度 卒業研究論文

題目 発言者識別を用いた対話型キャラクタの
セリフの違和感検出に関する研究

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 森 康汰

学籍番号 16024159

提出年月日 令和3年2月12日

目次

第 1 章	まえがき	1
第 2 章	違和感について	3
2.1	状況との不一致	3
2.2	個性との不一致	4
第 3 章	提案手法	5
3.1	提案手法 1	5
3.2	提案手法 2	6
3.2.1	語尾についての識別	6
3.2.2	口調についての識別	6
第 4 章	実験結果	11
4.1	データセット	11
4.2	評価尺度	11
4.3	提案手法 1 による実験結果と考察	12
4.4	提案手法 2 による実験結果と考察	13
4.5	口調についての識別器の閾値の検討	15
第 5 章	むすび	17
5.1	まとめ	17
5.2	今後の課題	18
5.2.1	語尾について	18
5.2.2	口調について	19
	謝辞	20
	参考文献	21

目次

1.1	提案システムのイメージ	2
3.1	2種類の提案手法の概観	8
3.2	発言者識別の概観	9
3.3	語尾に着目した違和感検出	9
3.4	口調に着目した違和感検出	10
4.1	閾値による正解率の推移	16
4.2	閾値による特異度の推移	16

表目次

4.1	評価実験用データセット	11
4.2	混同行列	12
4.3	SVM による発言者識別の再現率について	12
4.4	“クレヨンしんちゃん” における提案手法 1 の結果	12
4.5	“鬼滅の刃” における提案手法 1 の結果	12
4.6	提案手法 1 の違和感検出の結果	13
4.7	図 3.1 中の破線部の違和感検出の結果	13
4.8	“クレヨンしんちゃん” における提案手法 2 の結果	13
4.9	“鬼滅の刃” における提案手法 2 の結果	14
4.10	提案手法 2 の違和感検出の結果	14
4.11	類義語として抽出された失敗データの例	14
4.12	出現回数の閾値 t_n と類似度の閾値 t_s の検討	15

第1章

まえがき

現在の日本において、スマートフォンの普及に伴い若者を中心にソーシャルゲームが人気を博している。ソーシャルゲームの魅力としては、空いた時間で気軽に遊べること、基本無料で遊べるということ、飽きさせないよう短いスパンでイベントやストーリーをアップデートされることが挙げられ、その中でも短いスパンでのイベントやストーリーのアップデートされることがソーシャルゲームにおける一番の魅力であると考えられる。

しかしながら、短いスパンで作成されたキャラクターのセリフには、作成時間の短さ故にしばしばキャラクターに対して違和感のあるセリフが作成されてしまうという問題がある。実際にあった例として、“バンドリ”というソーシャルゲームに“今井リサ”というキャラクターがいるが、このキャラクターはストーリー上で不自然な形で家族に弟がいるという設定が追加され、ユーザーに混乱を招き炎上し結果としてゲーム運営会社が謝罪したという事例がある [1]。また、例として週刊誌のマンガがコミックになる時に加筆修正が多くなされる場合がある。加筆修正される内容についてだが、キャラクターのイラストへの加筆もあるが他のキャラクターへの呼称の修正や単語の変換に対する修正などが多い印象がある。これも先述したソーシャルゲーム同様、短いスパンで作成されたセリフであるため修正されるべき問題になっているのではないかと考える。

このような問題解決のため図 1.1 の様なシステムを構築することによって、キャラクターのセリフの違和感検出を機械が行うことになり、キャラクターのセリフ作成の時間削減や、セリフにおけるキャラクターの一貫性などの面で活躍が見込まれると考える。本研究では、あるキャラクターのセリフ中の語尾・口調などに対してテキスト解析を行い、セリフ中に違和感が存在するかを検出する手法について提案する。

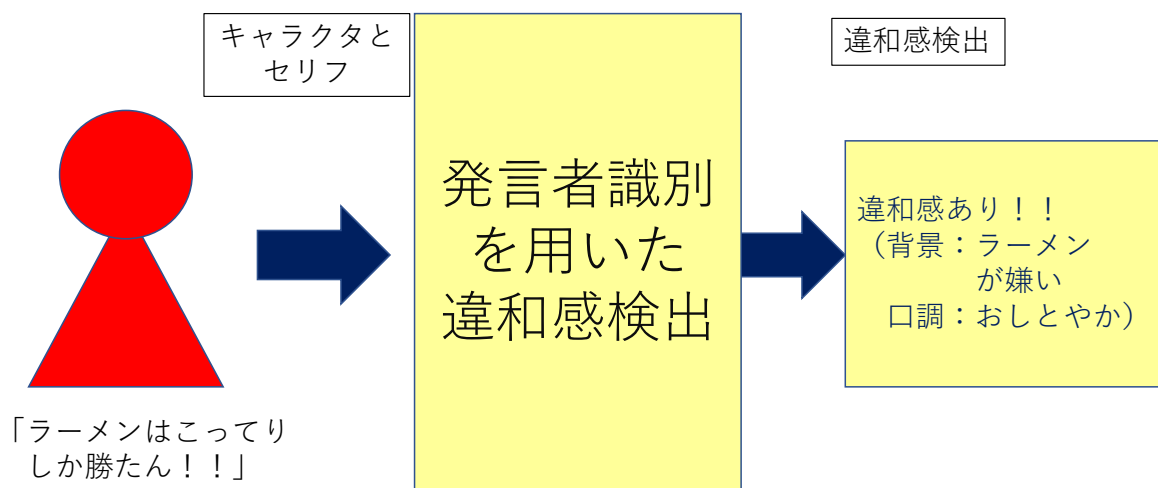


図 1.1 提案システムのイメージ

第2章

違和感について

ストーリーライターを補助するような先行研究として“セリフに基づく個性を考慮した話者の分散表現に関する考察”[2]などがすでに存在するが、セリフ中から違和感を検出する研究については不十分である。違和感には様々な種類のものがあるが、本研究における違和感を以下のように定義する。

2.1 状況との不一致

キャラクターが発言する際の状況とセリフの内容に対する考慮しなくてはならない違和感について以下の通り定義する。

対話相手との関係性

例. 日常的に立場が自分より上の方へ敬語で話している人が急にため口で話す

発言者の感情

例. 発言する瞬間に発言者がとても怒っている時に冷静なセリフが使用されている

対話相手の感情

例. 対話相手が怒っている時にふざけているようなセリフが使用されている

発言する場面の空気

例. シリアスな雰囲気の中で無責任なセリフが使用されている

状況との不一致について違和感が存在する場合に、読み手が感情移入しにくくなるという問題に繋がる可能性があり、取り除かれるべき違和感であると考えられる。

2.2 個性との不一致

キャラクターの個性とセリフの内容に対する考慮しなくてはならない違和感について以下の通り定義する。

キャラクターの語尾・口調

例. キャラクターを作り出す際に付与される話し方に関する設定との矛盾が存在する時

キャラクターの背景・過去

例. キャラクターを作り出す際に付与される過去の設定との矛盾が存在する時

個性との不一致について違和感が存在する場合に、キャラクターの個性が損なわれるなどの問題に繋がる可能性があり、取り除かれるべき違和感であると考えられる。本研究ではキャラクターの個性が顕著に表れ、定量的に評価を出しやすいため、上記中のキャラクターの語尾・口調について焦点を当てるものとする。

第3章

提案手法

キャラクタのセリフにおいて違和感を検出する手法として、ベーシックな発言者識別を使用する手法と違和感に着目した識別器と共に使用する手法の2つを提案する。本章では2種類の提案手法の概観を図3.1に示し、違和感検出の方法について詳述していく。

3.1 提案手法1

提案手法1(図3.1の上部)として、発言者識別を用いた違和感検出について提案する。その詳細は図3.2に示す。本研究では、Pythonのオープンソース機械学習ライブラリのscikit-learn[3]中のSVMを用いて違和感の有無について分類する。セリフの特徴ベクトルとしてBoW(Bag of Words)を使用する。変換された特徴ベクトルにSVMを使用し学習・予測を行う。BoWの処理については以下の通りである。

1. セリフを分かち書きによって単語に分解する
2. 各単語にインデックスが割り振られる
3. 文ごとに各単語の登場回数を数えインデックスを参照し並べる

ただし、その特徴ベクトルには、形態素解析ソフトMeCabを使用し、それらの単語の基本形を用いる。

また、提案手法1の違和感について、あるアニメーション作品のキャラクタについて学習させた発言者識別器を二次創作の同作品の同キャラクタに対して使用し、その結果として二次創作のキャラクタのセリフがアニメーションのキャラクタのセリフであると識別されなかった場合に違和感があると定義する。

3.2 提案手法 2

提案手法 2 はキャラクターの語尾に着目した識別器とキャラクターの口調に着目した識別器を用いた違和感検出について提案する。図 3.1 に記述されているフローチャートに従い、セリフを対象に語尾についての識別を行い、“違和感なし”と識別されたセリフに、口調についての識別を行う。その過程で“違和感あり”と一度でも識別されたセリフは“違和感あり”と予測され、“違和感なし”と予測されたセリフのみに提案手法 1 を使用し分類させるという手法である。

3.2.1 語尾についての識別

キャラクターの語尾を識別するために、語尾リストを用いる。理想として、この語尾リストにはあるキャラクターとそのキャラクターがよく使用する語尾のペアが複数、キャラクター分格納されている。本研究では簡易的に以下のような語尾リストを作成した。まず、様々なキャラクターがよく使用する語尾のみをウェブ上のニコニコ大百科 [4] からおよそ 200 件取得する。次に、本研究では特定のキャラクター c に限定してセリフの違和感検出を行うことを予め決めておき、200 件のデータに対して、そのキャラクター c の語尾であるか否かのラベルを著者が付与した。

具体的な違和感検出の処理は図 3.3 の通りである。まず、あるセリフ s に対して形態素解析を行い、以下の条件を満たす形態素を語尾とする。

- その形態素がそのセリフ s の最後の形態素で、且つ助動詞であるとき
- その形態素が最後から K 番目の形態素で、且つ助動詞であり、 $K + 1$ 番目から最後まで形態素がすべて記号であるとき

次に、そのセリフ s の語尾がキャラクター c の語尾リストの中に含まれており、且つその語尾がキャラクター c の語尾ではないとラベル付けされていた場合、そのセリフ s がキャラクター c のセリフとして違和感があると分類する。ただし、4 章の評価実験では語尾リストに含まれる 200 件は予め形態素解析のユーザ辞書として登録した。

3.2.2 口調についての識別

例えば、普段「お姉さん」とよく発言するキャラクターが、急に「お姉ちゃん」と発言したり、自分の事を「オラ」と発言するキャラクターが、急に「ぼく」と発言したりする時、我々はそれらに違和感を感じるであろう。このように、あるキャラクターの口調に違和感を覚えてしまう状況の 1 つとして、そのキャラクターが何度も使用する言葉があるにも関わらず、その言葉が使われていない場合が挙げられる。

そこで本研究では、あるキャラクターが発言したセリフ中で「置き換えた方がそのキャラクターのセリフとして自然である」という別の単語が認識された時、そのセリフはそのキャラクターのセリフとして違和感があると識別する手法を提案する。

具体的に、セリフ s がキャラクター c のセリフとして違和感があるか否かを識別する処理に

ついて説明する．まず，そのセリフ s に出現する単語群を W_i^s ($i = 1, 2, \dots$) とする．この時「キャラクター c のセリフとして，単語 W_i^s を別の単語 W に置き換えるべき」であるという状況は，その置き換え先の単語 W が以下の条件を満たしている時であるという仮説を立てた．

- 単語 W はキャラクター c が頻繁に使う単語である．
- 単語 W を単語 W_i^s と置き換えてもセリフの意味は変わらない．
(単語 W と単語 W_i^s とは意味的に類似する)

そこで本研究では，セリフ s の，キャラクター c のセリフとしての違和感の有無を示す $\text{hasJerk}(s, c)$ の定義式を以下の通り提案する．

$$\text{hasJerk}(s, c) = \begin{cases} 1 & (\exists i, \exists j, \text{sim}(W_i^s, W_j^c) \geq t_s \\ & \text{and } N(W_j^c) \geq t_n) \\ 0 & (\text{otherwise}) \end{cases}$$

ただし， W_j^c ($j = 1, 2, \dots$) は，キャラクター c の過去のセリフに出現した単語群， $N(W_j^c)$ ($j = 1, 2, \dots$) はその出現回数を表している．そして，キャラクター c の過去のセリフに頻出し，且つセリフ s の中のある単語 W_i^s の意味を変えないような単語 W_j^c が1つでも存在する時，単語 W_i^s は単語 W_j^c に置き換えた方が良く，セリフ s はキャラクター c のセリフとして違和感があると認識することになる．

本研究では，4章の評価実験の際に単語間の類似性の有無 $\text{sim}(W_i^s, W_j^c)$ について，類義語，上位語，下位語などがまとめられた概念辞書である WordNet[7] を使用する．また，類義語として抽出される単語には類似性があるものとする．使用例を図 3.4 に示す．

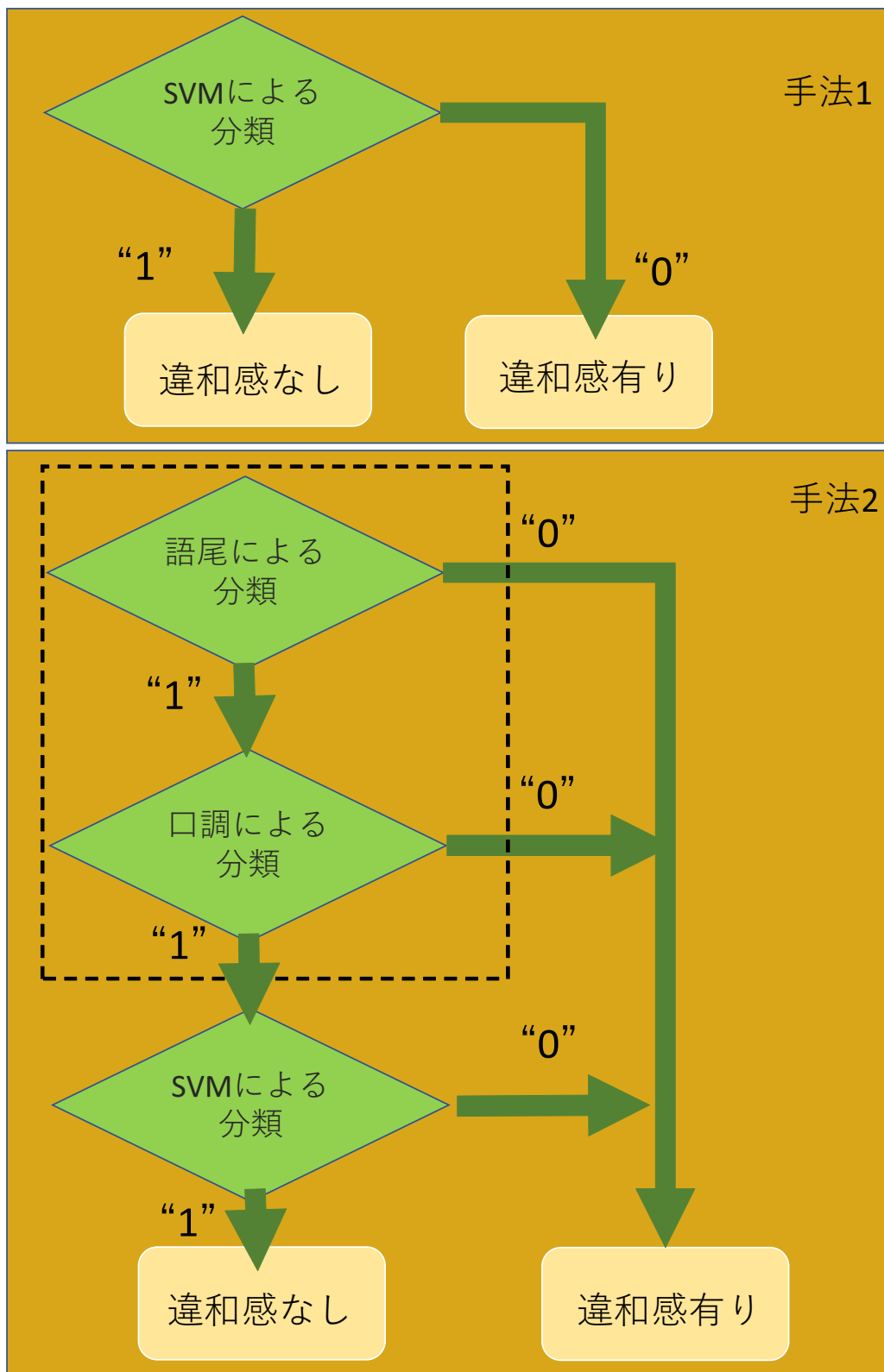


図 3.1 2種類の提案手法の概観

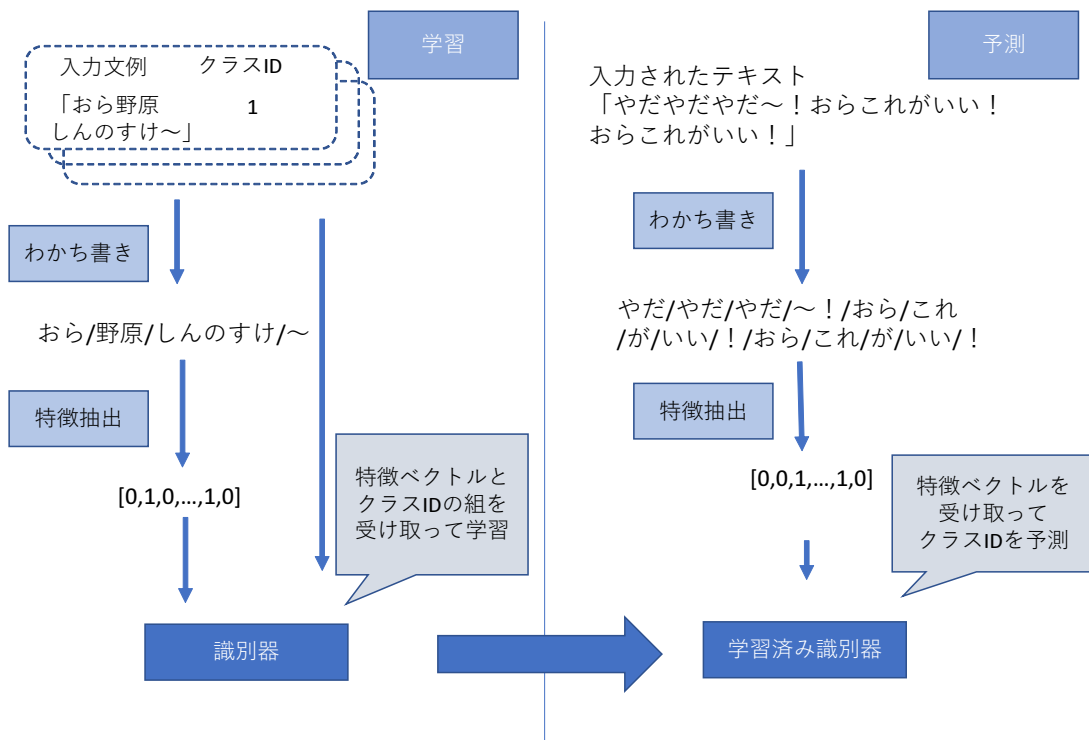


図 3.2 発言者識別の概観

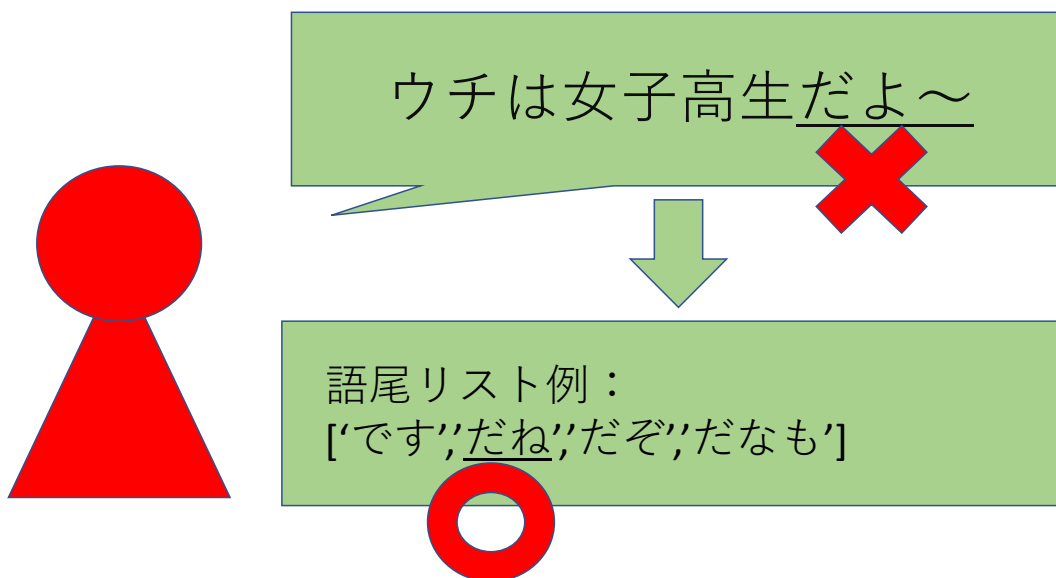


図 3.3 語尾に着目した違和感検出

入力されたセリフ(S)例：

「今日 も 化粧 完璧 だわ～」

w_1^S w_2^S w_3^S ... w_i^S

用意

キャラクタ(C)の口調リスト例：

["君", "だよ～ん", "ケータイ", "メイク"...]

w_1^C w_2^C w_3^C w_4^C ... w_j^C

置き換え可能か判別：

$$hasJerk(s, c) = \begin{cases} 1 & (\exists i, \exists j, sim(W_i^S, W_j^C) \geq t_s \\ & \text{and } N(W_j^C) \geq t_n) \\ 0 & (otherwise) \end{cases}$$

置き換え可能例：

(["今日", "化粧", "完璧"], "メイク")

$sim(\text{"化粧"}, \text{"メイク"}) \geq t_s$

$\text{and } N(\text{"メイク"}) \geq t_n$



違和感あり！！

図 3.4 口調に着目した違和感検出

第4章

実験結果

本章では2種類の提案手法にて違和感検出を行い、それぞれの提案手法が優れている度合いについて評価し考察する。

4.1 データセット

本実験では、学習データとしてNETFLIX[5]にて公開されているアニメーション作品の字幕データを使用し、特定のキャラクタか、それ以外のキャラクタかでラベル付けを行い学習モデルを作成する。本実験では、表4.1に示す2つのアニメーション作品の字幕データを使用する。また、真偽の割合に対して違和感のないセリフを真とする。

表 4.1 評価実験用データセット

作品	発言者	ジャンル	真偽の割合
クレヨンしんちゃん	野原しんのすけ	日常	3:2
鬼滅の刃	竈門炭治郎	アクション	9:1

テストデータとして二次創作の小説まとめサイト [6] から学習データに使用した作品の小説を使用する。データ単位について、小説中で改行が発生するまでを1データとする。また、二次創作の小説にはアニメーションで発言されないようなセリフ（違和感があるセリフ）があるものとし、違和感の有無をラベル付けしたデータを実験に使用する。

4.2 評価尺度

本実験では、表4.2に示す混同行列を算出した後、正解率、特異度を用いて、予測性能を評価する。正解率について、違和感がないセリフに対し違和感がないと、違和感があるセリフに対し違和感があると識別出来た割合を示し、発言者識別と違和感識別の総合評価値として使用する。特異度について、実際に違和感があったセリフの中で違和感を再現できた割合を示す。

表 4.2 混同行列

	違和感なしと予測	違和感ありと予測
違和感なしデータ	TP	FN
違和感ありデータ	FP	TN

$$\text{正解率} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{再現率} = TP / (TP + FN)$$

$$\text{特異度} = TN / (TN + FP)$$

4.3 提案手法 1 による実験結果と考察

まず、提案手法 1 で作成した違和感検出システムの精度の検証として、アニメーション中で使用されている特定のキャラクタのセリフのみを提案手法 1 に対し使用する。その結果を表 4.3 に示す。次に提案手法 1 を用いて二次創作のセリフに対し使用した結果を表 4.6 に示す。

表 4.3 SVM による発言者識別の再現率について

評価尺度	クレヨンしんちゃん	鬼滅の刃	平均値
再現率	0.223	0.336	0.279

表 4.4 “クレヨンしんちゃん” における提案手法 1 の結果

	違和感なしと予測	違和感ありと予測
違和感なしデータ	38	51
違和感ありデータ	21	33

表 4.5 “鬼滅の刃” における提案手法 1 の結果

	違和感なしと予測	違和感ありと予測
違和感なしデータ	6	46
違和感ありデータ	0	7

表 4.6 提案手法 1 の違和感検出の結果

評価尺度	クレヨンしんちゃん	鬼滅の刃	平均値
正解率	0.496	0.220	0.358
特異度	0.611	1.000	0.805

SVM の再現率について、表 4.3 から“クレヨンしんちゃん”、“鬼滅の刃”ともに再現率が低い値となり、解析されたセリフが、そのキャラクターのセリフであると識別出来なかったため違うキャラクターのセリフであると検出される割合が多いことが分かる。これについては特徴ベクトルの作り方や識別器自体に問題があるのではないかと考えられる。二次創作に対する SVM による違和感検出の精度として、表 4.6 から、正解率では“クレヨンしんちゃん”の方が高く、特異度では“鬼滅の刃”の方が高いことが分かる。“鬼滅の刃”より“クレヨンしんちゃん”の正解率が高いことについて、二次創作中の違和感のないセリフがアニメーション中のキャラクターのセリフであると識別されている割合が“鬼滅の刃”より“クレヨンしんちゃん”の方が高くなっている。また、“鬼滅の刃”の識別結果が全体的に違和感ありの分類に偏っていることから、正解率が低くなっているのではないかと考えられる。“クレヨンしんちゃん”より“鬼滅の刃”の特異度が高いことについては、“鬼滅の刃”の二次創作中に違和感が含まれるセリフが少ないことや、違和感ありと検出されることが多いことから、違和感のあるセリフをすべて検出できたと言える。

4.4 提案手法 2 による実験結果と考察

図 3.1 中の破線部に当たる語尾・口調による違和感検出の結果が表 4.7 の通りである。次に図 3.1 中の下部に当たる語尾・口調と SVM について違和感検出の結果が表 4.10 の通りである。

表 4.7 図 3.1 中の破線部の違和感検出の結果

作品	実際に違和感あり / 分類器が違和感あり
クレヨンしんちゃん	0.538 (21/39)
鬼滅の刃	0.076 (1/13)

表 4.8 “クレヨンしんちゃん”における提案手法 2 の結果

	違和感なしと予測	違和感ありと予測
違和感なしデータ	25	64
違和感ありデータ	6	48

表 4.9 “鬼滅の刃”における提案手法2の結果

	違和感なしと予測	違和感ありと予測
違和感なしデータ	6	46
違和感ありデータ	0	7

表 4.10 提案手法2の違和感検出の結果

評価尺度	クレヨンしんちゃん	鬼滅の刃	平均値
正解率	0.510	0.220	0.365
特異度	0.888	1.000	0.944

表 4.10 の 2 つの作品の実験結果の平均値が提案手法 1 の値よりも高いため、提案手法 2 の語尾・口調について違和感検出を行うことで精度が向上したと考えられる。語尾の識別について、“野原しんのすけ”のセリフに対してはキャラクタの語尾がはっきりしているため違和感が検出しやすく判別結果もかなり良いものとなった。“竈門炭治郎”のセリフに対しては語尾に個性が出てこないため有効ではなく、口調による判断がほとんどであった。

語尾の識別器・口調の識別器について、“野原しんのすけ”のセリフに対しては正解率を一定の数値に保ったまま、違和感があるセリフに対して正しく予測することができたため、語尾の識別器・口調の識別器は良い精度が出ていると考えられる。“竈門炭治郎”のセリフに対しては、提案手法 1 と同じ結果となった。これは“竈門炭治郎”というキャラクタが語尾・口調に個性が強く出てこないことが原因ではないかと考える。提案手法 2 で抽出できなかった例を表 4.11 に示す。

表 4.11 類義語として抽出された失敗データの例

作品	抽出された単語	類義語
クレヨンしんちゃん	暗い	重い
クレヨンしんちゃん	助け	世話
鬼滅の刃	仲間	相手

提案手法 2 の口調についての違和感の定義式中の t_s 値と t_n 値について、本研究では t_s については“WordNet に類義語が存在するかどうか”として、 t_n 値については $t_n=1$ として実験を行った。その結果として取得した失敗データの例を表 4.11 に示す。違和感の誤検出を防ぐためには、 t_s 値と t_n 値のパラメータについて閾値を算出することが重要ではないかと考える。

4.5 口調についての識別器の閾値の検討

3.2.2 節で定義した過去に発言した単語の出現回数 $N(W_j^c)$ について、本実験では一回でも出現している単語を置き換え対象として実験を行った。また、単語の類似性の有無 $\text{sim}(W_i^s, W_j^c)$ について、WordNet にて類義語として取得される単語を置き換え対象として実験を行った。ここで置き換え対象となるセリフの誤検出を防ぐために単語の出現回数と単語間の類似度の閾値について検討を行う。単語の出現回数 $N(W_j^c)$ の閾値 t_n について、 $t_n \in \{1, 2, 3, 4\}$ でパラメータを動かし、 t_n 値以上の出現回数の置き換え先単語を含むセリフを違和感が存在するセリフとして検出する。単語間の類似性 $\text{sim}(W_i^s, W_j^c)$ を方法として、WordNet にて類義語として取得されるか否かに基づく $\text{sim} \in \{0, 1\}$ から、単語の分散表現の fastText [8] を用いた $\text{sim} \in [0.0, 1.0]$ に改良する。また学習データには、Wikipedia のデータを使用する。単語間の類似性 $\text{sim}(W_i^s, W_j^c)$ の閾値 t_s について、値を 0.1 ずつ増やしていき t_s より大きい類似度を含むセリフを違和感が存在するセリフとして検出する。閾値に関する実験の結果を表 4.12、図 4.1、図 4.2 に示す。

表 4.12 出現回数の閾値 t_n と類似度の閾値 t_s の検討

		$t_s=0.1$	$t_s=0.2$	$t_s=0.3$	$t_s=0.4$	$t_s=0.5$
$t_n=1$	正解率	0.510	0.517	0.517	0.524	0.517
	特異度	0.888	0.888	0.870	0.851	0.777
$t_n=2$	正解率	0.517	0.517	0.524	0.531	0.538
	特異度	0.814	0.814	0.814	0.796	0.777
$t_n=3$	正解率	0.524	0.524	0.531	0.531	0.538
	特異度	0.814	0.814	0.814	0.796	0.777
$t_n=4$	正解率	0.531	0.531	0.531	0.531	0.538
	特異度	0.814	0.814	0.814	0.796	0.777

結果から、まず出現回数 $N(W_j^c)$ の閾値 t_n について、閾値を大きくしていくにつれ正解率の値が高くなっていき特異度の値が低くなっていくことが分かる。次に単語間の類似性 $\text{sim}(W_i^s, W_j^c)$ の閾値 t_s について、閾値を大きくしていくにつれ特異度の値が低くなっていることが分かる。考察として、単語の出現回数の閾値・単語間の類似度の閾値を増加することによって違和感の誤検出が減り正解率が上がったのではないかと考える。 $t_s=0.5$ 、 $t_n \geq 2$ の時正解率が 0.538 と最大となったが、これは口調に関して違和感検出をほぼ行わず、語尾についての識別器と SVM での精度となっている。また本実験では WordNet から類義語として抽出された単語間で類似度を算出していたが、WordNet から類義語として抽出される単語数が多く、その中に類似度の高い単語が含まれており誤検出されてしまったのではないかと考える。

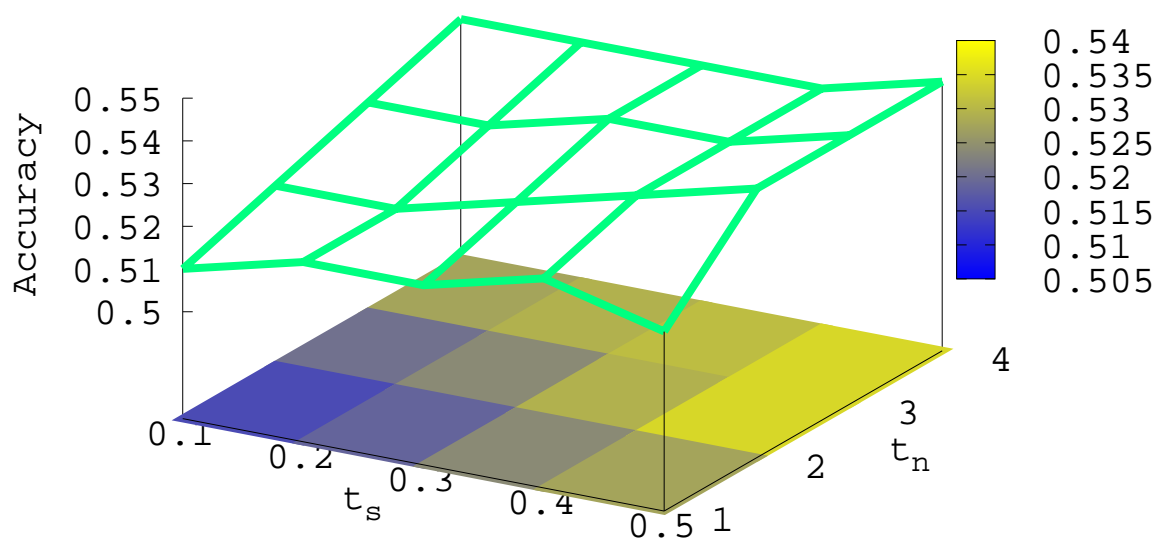


図 4.1 閾値による正解率の推移

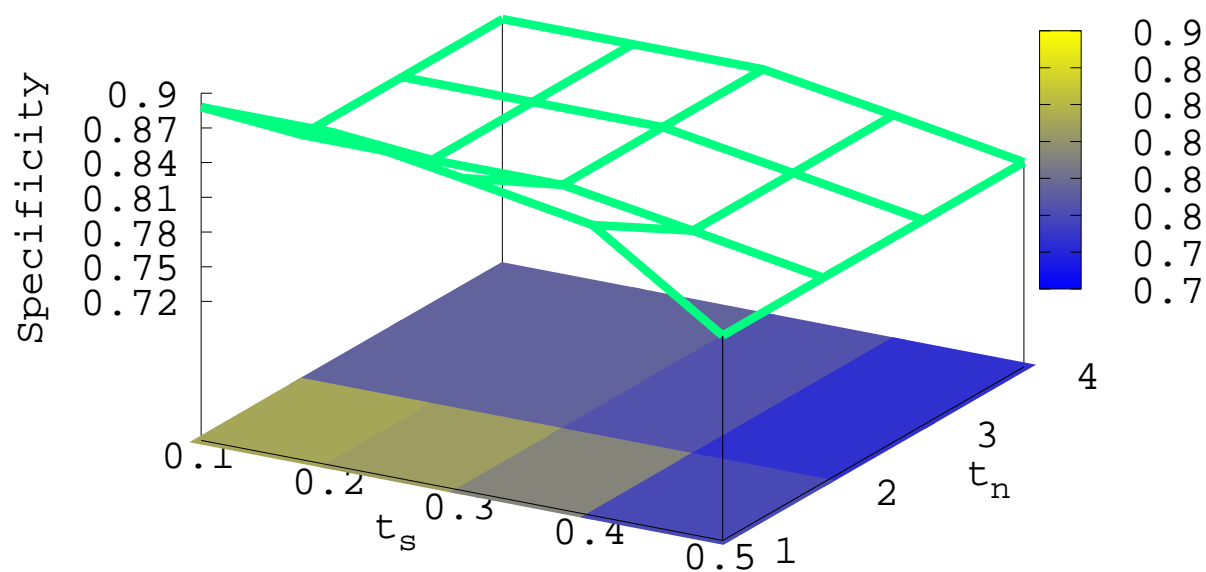


図 4.2 閾値による特異度の推移

第 5 章

むすび

本章では、本研究におけるまとめと提案手法における今後の課題について詳述していく。

5.1 まとめ

本研究では発言者識別を用いたキャラクターのセリフの違和感検出に関する研究を行ってきた。本研究は、先行研究の話者推定とは異なり、違和感検出を目的としているため、評価尺度の正解率・特異度の値の変化について着目した。二次創作の小説をテストデータとして違和感検出を行ってきたが、読み手のキャラクター像と二次創作の作者のキャラクター像の間にギャップを感じ、それが違和感の元となったのではないかと考えられる。表 4.6, 表 4.10 から、提案手法 1, 提案手法 2 共に特異度の値が高いことがわかり、また提案手法 1 と提案手法 2 を比較した時、正解率をほぼ一定にしたまま特異度が上がっていることが分かる。提案手法 2 で行っている違和感検出により提案手法 1 より違和感ありと識別される割合が多くなるが、正解率の値がほぼ一定ということから違和感検出の精度が向上したことを示していると考えられる。しかし、本研究で使用した作品 2 種共に正解率の値が低く、現時点では実用段階には至っていないと考えられ、提案手法 1 で本研究ではベーシックな SVM による分類手法を使用したのが、他のテキスト分類システムを使用することによって正解率の値が高くなるのではないかと考える。

また、本研究では違和感について特にキャラクターの語尾・口調に着目したモデルについて提案したが、“野原しんのすけ”のような、セリフに語尾・口調の個性が出てくるキャラクターに対しては良好的に違和感を検出できるが、“竈門炭治郎”のような個性が出てこないキャラクターに対してはあまり良い結果にならなかったため、他の違和感について着目した識別器の作成も検討の余地があると考えられる。

本研究の現状の社会的貢献は、違和感に着目した識別器を作成したことによるキャラクターのセリフのクオリティの向上である。1 章でも述べたが、短いスパンで作成されたキャラクターのセリフには違和感が含まれていることがしばしば存在するという問題がある。現状では、エンタメというジャンルにおいて更新するスパンの短さとクオリティの高さが求められていると考え、クオリティの向上のために違和感に着目した識別器の作成は問題解決のアプローチとして有効ではないかと考える。また、例として、昨今の就職事情では、企業の人事担当が就活生の

情報を得るために就活生の SNS のアカウントをチェックし合否に反映するという話がある。その際に SNS のアカウント特定方法として発言者識別を用いることにより、SNS のアカウントの発言が就活生の発言に近いが、また就活生の面接時のイメージと SNS 上での発言に差があるかを効率よく確認できるため違和感検出は有効ではないかと考え、社会的に貢献できていると考える。

本研究の現状の技術的貢献は、自然言語処理という分野で違和感について定義し、違和感という尺度からテキスト分類を行えたことである。違和感について、画像に対する視覚的違和感の研究が多く行われており、自然言語処理中のセリフにおける違和感に対する研究については不十分であると考え。本研究では使用した定義式から発言者識別と違和感検出の精度向上という面で貢献できているのではないかと考えられる。

5.2 今後の課題

今後の課題として、提案手法 2 の改良案について考察する。

5.2.1 語尾について

語尾についての識別器の判断基準が“入力されたセリフに語尾が入っているか”と、“入力されたセリフ中の語尾が識別したキャラクターのものか”で判断していた。本研究では判断基準として使用しなかったが、“入力されたセリフに語尾が無かった場合、語尾を追加できるか”という観点からも違和感が発生すると考え、入力されたセリフに語尾になりえる単語を追加し、違和感の識別を行うことができれば、より実用的な語尾の違和感検出が可能となると考える。

また、語尾データについて、インターネット上の語尾リストから特定のキャラクターの語尾か否かのラベルを付与し、実験に使用しているが、語尾リストが更新されないため、新しい語尾に対応できないことや、キャラクターの語尾について裏付けが取れないという問題点が存在する。この問題の解決案として、学習データ中から語尾を自動で追加する方法について検討をする必要があると考える。

5.2.2 口調について

本実験では口調の置き換え語に対して品詞全てではなく、一部の品詞のみで行っていたが、今後は品詞に対する検討を行い、抽出される単語から個性や違和感が検出されるか研究を進めていこうと考えている。

本研究では単語間の類似性について WordNet から取得される類義語には類似性があるものとして使用していたが、その結果として以下のような単語の置き換えが誤検出された。

「難しい」「よく」 → 「大変」
「すごく」「きれい」 → 「いい」

上記から置き換え先の単語のような“広義的に類似である単語”が存在することが分かった。この問題の解決案として、置き換え先の単語が広義的に使われている単語であるかどうかを判断することによって誤検出が減り正解率が上がるのではないかと考えられる。

また、本システムでは図 3.1 のフローチャートに従い語尾についての識別器、口調についての識別器、SVM の順で識別処理を行った。そのため、語尾についての識別器・口調についての識別器の段階で違和感を誤検出した場合、正解率が下がるという問題がある。この問題の解決案として、並列処理について検討する。システムについては入力されたセリフをそのまま各識別器に使用し、その結果から多く出力されたものを採用するという構成である。本実験では識別器の数が少ないという事や、語尾についての識別器・口調についての識別器がそれぞれ一定の条件下でしか分類を行わないため並列処理を行わなかったが、今後複数の識別器を使用する際や各識別器の精度が向上した際に並列処理を実装しようとする。

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にも深く感謝の意を表します。

参考文献

- [1] AnimeAnime.jp, <https://animeanime.jp/article/2019/10/23/49148.html> (2020/12/18 参照).
- [2] 佐藤 幸一, 福田 清人, 森 直樹, 松本 啓之亮, “台詞に基づく個性を考慮した話者の分散表現に関する考察”, 言語処理学会 第 24 回年次大会 発表論文集, pp.1280–1283 (2018). https://anlp.jp/proceedings/annual_meeting/2018/pdf_dir/C7-4.pdf (2020/12/18 参照).
- [3] scikit-learn, <https://scikit-learn.org/stable/> (2020/12/18 参照).
- [4] ニコニコ大百科, <https://dic.nicovideo.jp/a/%E8%AA%9E%E5%B0%BE%E3%81%AE%E4%B8%80%E8%A6%A7> (2020/12/18 参照).
- [5] NETFLIX, <https://www.netflix.com/> (2020/12/18 参照).
- [6] 二次創作の小説まとめ, エレファント速報, <http://elephant.2chblog.jp/> (2020/12/18 参照).
- [7] 日本語 WordNet, <http://compling.hss.ntu.edu.sg/wnja/> (2020/12/18 参照).
- [8] fastText, <https://fasttext.cc/> (2021/02/01 参照).