音声合成の発話スタイルを切り替えるシステムの設計

2018SE071 太田侑雅　2018SE074 坂瞭太郎
指導教員：野呂昌満

1　はじめに

テキスト音声合成は，入力された任意のテキストから音声を自動生成する技術である．近年ではただ文章を読み上げるだけでなく，喜びや悲しみといった感情を 1 つ選択し，その感情で文章を読み上げるツールが数多く存在する．これらのツールは，文章内容と音声合成による感情表現を機械的に一致させることができない． 山岸らの研究 [1] では HMM(隠れマルコフモデル) 音声合成の決定木によるコンテキストクラスタリングで発話スタイルを制御し，文の発話スタイル切り替えを可能とした．

本研究では, 文章をフレーズ単位に分割し, フレーズ内容に合致した感情音声で読み上げるシステムを提案する．このシステムはフレーズ分割, 感情分類器, 音声出力ソフトウェアの 3 つの構成要素で構築する. 本研究では感情分類器にナイーブベイズ分類, 音声出力ソフトウェアに HMM 音声合成といった要素技術を利用し既存の構成とした.

本研究の目的はアーキテクチャを設計し, その構成要素となる要素技術は既存のものを利用し, それらを統合することでシステムを構築することである.

本研究の技術課題を以下に示す.

課題 1　提案システムのアーキテクチャの設計

課題 2　要素技術の確認と既存システムの再利用による提案システムの設計（要素技術の統合）

課題 3　実データを用いた提案システムの動作確認

この時, 技術統合においての重要課題は以下の 2 点である.

課題 2.1　感情分類器の教師データとなる感情コーパスの構築

課題 2.2　ナイーブベイズ感情分類器の妥当な精度の確立

2　既存研究ならびに関連技術

2.1　ナイーブベイズ分類器

山本ら [2] の研究では感情コーパスを構築する目的で, 感情分類手法としてナイーブベイズ分類を提案している．ナイーブベイズ分類とは過去の事例をもとに未知の文書があらかじめ与えられているどのカテゴリに属するかを決定する分類手法である.

2.2　BoW（Bag of Words）

BoW(Bag of words) とは自然言語処理において人間が日常的に使用する自然言語で記述されたテキストをベクトルで表現する手法である.

3　提案システムの設計

本研究で提案するシステムのアーキテクチャを図 1 を示す.

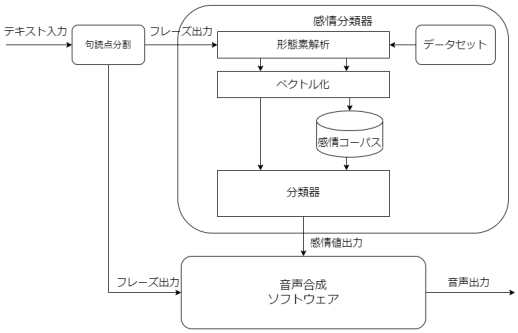

図 1　提案システムのアーキテクチャ

本アーキテクチャは本質的に言語処理を行うものであると考え, 基本的に言語処理系のアーキテクチャに倣った. 感情表現は形態素レベルのもので十分であると考えたので, 構文解析部は省略する構成となり図 1 となった.

このアーキテクチャの動的挙動は以下のとおりである.

(1) 句読点分割された感情ラベル付きテキストデータセットを形態素解析をし, それをベクトル化した感情コーパスを構築する.

(2) ユーザによって入力されたテキストを句読点分割する.

(3) フレーズ毎に形態素解析, ベクトル化を行う.

(4) 構築された感情コーパスを用いて, フレーズ毎に感情分類する.

(5) 感情分類器で感情分類された感情値を音声合成ソフトウェアに出力し, 感情値に対応する音響モデルでフレーズの音声合成を行い, 音声出力する.

3.1　感情コーパス

感情コーパスとは感情分類する際に指標となるデータである．用いるテキストデータは句読点分割されたものとする．コーパスの元となるテキストデータセットは文学作品を一部抜粋したものや大学共同利用機関法人人間文化研究機構及び独立行政法人情報通信研究機構 [3] が所有する「日本語話し言葉コーパス」から無作為に抽出した文章を利用する．これらの文章は感情がラベル付けされていないので，本研究では人手で文章を「喜び」「悲しみ」「怒り」「平常」の 4 感情に分類する．その評価基準として日本語感情表現辞書 [4] を用いる．日本語感情表現辞書を用いる利点として，感情語の感情方向に定量化を計れることである．抽出する文に対してすべて人力でカテゴリ分類するには評

価者の主観が多く織り込まれてしまう. 従って感情表現辞書を用いることによって評価の分散が減少すると考える.

文中に [4] の語が含まれている場合，その語句の感情値を文章にラベリングする．文中に複数の感情語が含まれている場合, 感情語の総数が多い方の感情にラベリングする．文中の複数感情の語句総数が等しい場合，主観的評価により文章をラベリングする．文中に [4] の語が含まれていない場合，主観的評価により文章をラベリングする. このときラベリングの主観性を減少させる目的で, その必要最低限の人数である我々研究者二人でラベリングを行う.

3.2 句読点分割

入力テキストの句読点分割は Python のライブラリにある findall 関数を用いる.

3.3 感情分類器

本研究では山本らの研究 [2] を参考にし, 感情分類手法にナイーブベイズ分類を用いる．ナイーブベイズ分類による感情分類器の設計は以下のようになる.

(1) テキストデータの形態素解析

(2) BoW(Bag-of-Words)

(3) (2) でベクトル化したデータをナイーブベイズ分類

3.3.1 テキストデータの形態素解析

テキストデータの形態素解析には Python の形態素解析ツール Janome を用いる．形態素解析の対象とする品詞のパターンを以下に示す.

(1) 動詞，形容詞，助動詞

(2) 動詞，形容詞，助動詞，名詞

(3) 全ての品詞

(1) の 3 種類の品詞に着目した理由は，思考やそれに伴った行動を示す言葉は形容詞や動詞であり，感情との関連性が高いと考えるからである．助動詞は形容詞や動詞を否定する「ない」を考慮するからである．(2) は (1) の品詞に加えて，文章の構成において一般的に最も出現回数の多い品詞の名詞を追加することで，感情分類器の判断材料を増やしている.

3.3.2 BoW(Bag-of-Words)

3.2.1 で形態素解析して得られた単語の辞書を作成する．辞書に登録された単語の種類を n とし，各テキストデータの単語の出現回数を n 次元ベクトルで表す.

3.3.3 ベクトル化したデータをナイーブベイズ分類

本研究では形態素解析レベルの感情表現を行うので, 単語やフレーズの特徴量を独立に考えるナイーブベイズ分類を用いる.

3.4 音声ソフトウェア

本研究で用いる音声ソフトウェアは徳田ら [5] が開発した,「喜び」「悲しみ」「怒り」「平常」の 4 感情の音響モデルを有する，HMM 日本語テキスト音声合成ソフトウェア「Open-JTalk」とする.

4 要素技術の統合

本研究では第 3 章で紹介した要素技術を統合することで，設計したアーキテクチャの実現が可能となる．要素技術の統合の重要点は以下の 3 つである.

- 句読点分割＋感情分類器
- BoW(入力テキスト) ＋感情コーパス＋ナイーブベイズ分類
- 感情分類器＋ Open-JTalk

句読点分割と感情分類器の統合は, テキストファイルをフレーズ数に合わせて作成することで実現し, フレーズ毎に感情分類できるようになる.

入力テキストと感情コーパスとナイーブベイズ分類の統合は, 入力テキストのベクトルの次元を, 感情コーパスの単語辞書の単語種類に合わせることで実現し, 入力テキストをナイーブベイズ分類を用いて感情分類できるようになる.

感情分類器と Open-JTalk の統合は, 文の冒頭フレーズから順に感情分類結果を出力することで実現し, フレーズ内容に合致した感情音声で読み上げるようになる.

5 精度検証の方法及び検証結果

感情コーパスのテキストデータを学習データとテストデータに分割し，ナイーブベイズ分類で学習，識別を行う．感情分類器の精度検証は K = 10 とした K 分割交差検証を用い, 以下の方法で行った. 方法 1 では精度が低い結果となったので方法 2 を用意した.

方法１: 感情コーパス構築に用いるデータセットは評価者二人の感情値が一致したもののみを扱い検証.

方法２: 評価者がそれぞれラベリングしたデータを足し合わせ, 更に新たなデータを追加したデータセットで検証.

以上の方法で検証を行った結果を以下に示す.

方法１の結果: 方法 1 の検証結果を表 1 に示す. 形態素解析の対象品詞が「動詞、形容詞、助動詞」の場合, 検証結果が他 2 つのパターンのものよりも高い精度を示したが,「動詞、形容詞、助動詞」を用いた場合でも精度は 70 %にも達せず，全体的に精度が低い結果となった.

方法２の結果: 方法 2 では表 2 より，データを 3523 個から 9424 個に加増させた．データ体系は, 評価者がそれぞれラベリングしたデータを足し合わせているので, 2 人の評価が一致していないものも含んでいる．増強

したデータを用いて感情分類精度検証をしたものが表3である．表3より，全体的に精度が向上したことがわかる．対象品詞が「動詞，形容詞，助動詞」の場合，その精度は約67 %を示し，他2つにおいても80 %を超える精度を示した．

以上の結果から，本研究で提案するシステムで用いる感情コーパスは最も検証精度の高かった，全ての品詞が分析対象の方法2のデータセットとする．

K分割交差検証 / 解析品詞の種類	単語の種類	平均正解率 (K=10)	正解率の標準偏差
動詞、形容詞、助動詞	966	約0.680	約0.049
動詞、形容詞、助動詞、名詞	3953	約0.504	約0.111
全ての品詞	4431	約0.520	約0.128

表1　評価値が一致したもののみの精度検証

感情	フレーズ数	フレーズ数
喜び	305	1516
悲しみ	459	1728
怒り	261	1112
平常	2498	5068
合計	3523	9424

表2　精度検証に用いたデータの数

K分割交差検証 / 解析品詞の種類	単語の種類	平均正解率 (K=10)	正解率の標準偏差
動詞、形容詞、助動詞	1230	約0.670	約0.040
動詞、形容詞、助動詞、名詞	4720	約0.817	約0.059
全ての品詞	5255	約0.837	約0.072

表3　増強されたデータセットを用いた精度検証

6　研究成果

実データを用いた提案システムの動作確認とその評価を以下に示す．

成果

(1)　出力される音声ファイルがフレーズ毎に連続して再生される

(2)　分類された感情の音響モデルで音声出力できる

問題

(1)　フレーズごとの音のつながり，特に「喜び」の音声と他の感情音声とのつながりに違和感を感じる

(2)　同じ意味でも漢字とひらがなでは分類結果が違う場合がある

(3)　感情カテゴリの種類が少ない

(4)　各感情の強弱に差がない

7　考察

7.1　課題1に対する考察

我々が思索したようにフレーズ分割，感情分類器，音声合成ソフトウェアを統合することで提案システムを実現することができた．このことから本研究の目的に対して有効なアーキテクチャが提案されたと考える．

7.2　課題2.1並びに課題2.2に対する考察

感情分類精度の向上において重要な要素はデータと品詞であると考え，それらについて考察していく．

7.2.1　感情分類器で用いた品詞について

表1と表3から，3つの品詞による感情分類方法は高い精度を得られなかったことがわかる．その原因として，単語の総数や種類，助詞などの数が関係していると考える．

表1では3つの品詞による分類精度が最も高いが，実データを用いた動作確認をしたところ，ほとんどのフレーズが平常に分類された．これは全体の約70 %のデータが平常カテゴリに属していることと，コーパスの辞書に登録する単語の種類が他の2つの分類方法に比べて少ないことが原因だと考える．

表3ではすべての品詞を用いたものが最も高い分類精度を示した．過学習が発生している可能性はあるが，この結果は単純に単語の総数や種類が最も多いことが起因していると考える．

7.2.2　感情分類器で用いたデータについて

対照実験を行う目的で表4を作成し，用いたデータについて考察する．

表1は，両者のラベリングが一致したものをデータセットとし精度検証を行った結果である．

表4は，表1で用いたデータセットを評価者のラベリングの一致不一致問わず，ラベリングしたデータを合成したデータセットで精度検証を行った結果である．

K分割交差検証 / 解析品詞の種類	単語の種類	平均正解率 (K=10)	正解率の標準偏差
動詞、形容詞、助動詞	981	約0.784	約0.016
動詞、形容詞、助動詞、名詞	4007	約0.854	約0.037
全ての品詞	4487	約0.860	約0.046

表4　評価値が不一致のものも含めた精度検証

表1と表4を比較すると，精度検証にかなりの差があることがわかる．その差は，一致不一致を考慮する場合と，しない場合において優劣性が一概に言えない．その考察として以下の3つが挙げられる．

(i)　総単語数の違い

品詞の考察と同様に総単語数が増えた影響と考える．

(ii)　感情値の増幅

評価の一致不一致に関わらず，二人が評価したデータセットをすべて用いることで．感情値に差をつけたからではないかと考える．本研究では各フレーズに対して二人がそれぞれカテゴリ分類を行った．合致したものだけをデータセットとした場合，フレーズの感情

値は一方向に定まる. 合致させていないものも含めてデータセットとした場合, 各フレーズに最大 2 方向の感情値を持つことにはなるが, 1 方向の場合が他の感情値に比べて 2 倍になるため相対的に精度が上がったのではないかと考える.

(iii) 過学習発生の可能性

本研究で作成したデータセットでは, 過学習が起こっている可能性があると考える. 評価が一致していないものも含めたデータセットには, 評価者が同じ分類した場合にのみ、そのフレーズのベクトルの強度を上げるために同文が複数入っている. この場合,K 分割交差検証で分割される訓練データとテストデータそれぞれには, 同じフレーズが含まれている可能性がある. それにより過学習が発生してしまっている可能性がある.

7.3 課題 3 に対する考察

第 6 章で挙げられた問題についての考察を述べていく.

(i) フレーズ間の言語特徴量と音響特徴量の関係を独立に考えているからである. 本研究では音声出力システムとして HMM 音声合成システムの Open-JTalk を用いた.Open-Jtalk の音声は音の抑揚はあるものの，フレーズ間の音声のつながりや単語の発音に違和感を感じることがあった. そこで, 本研究で扱われた音声出力システムを DNN 音声合成システムに置き換えることで改善するのではないかと考える.DNN 音声合成システムとは言語特徴量と音響特徴量の関係を多層の NN で表現し，音声出力するものである．DNN 音声合成は HMM 音声合成よりも学習データを効率よく利用し, 高精度に音響特徴量を予測することが可能となり, 合成音声の自然性が高い特徴がある.

(ii) 形態素解析のみで言語解析を行っているからである. 同じ意味を持つ単語において, 感情分類を行ったとき, 異なる分類結果が得られるときがあった. 同じ意味の単語でも「漢字」「ひらがな」「カタカナ」などの表記ゆれがある. 本研究では形態素解析のみでの言語解析を行うので, そのような表記ゆれに対応することができず分類結果に差が生じてしまう. これを解消するには, 形態素解析された単語に対し表記統制辞書 [6] などを用意することによって解消できるのではないかと考える.

(iii) 基本的な感情は「喜び, 悲しみ, 怒り, 平常」の 4 つで構成されていると考えたからである. 加えて Open-JTalk が所有する感情音響モデルがその 4 感情の音響モデルしかなかったからである. 分類先が 4 つの影響で, 多くの感情語において感情の偏りが見られ，感情分類の精度に歪みが生じた可能性がある．従って感情の種類を増やすことで，文により適切な感情を割り当てることができると考える．感情音声に関しては (i) 同様，多様な感情音声を持ち合わせている，DNN 音声合成システムを用いることで解決できると考える.

(iv) ナイーブベイズ分類ではフレーズの特徴量が独立であり, 互いに相関がないからである. 従ってフレーズ間の関係性の考慮や感情の強弱を作ることができなかった．そこで「tf-idf」や「WordNet-affect」を用いることによって豊かな感情表現が可能であると考える.

tf-idf を用いることで，文書に含まれる単語の重要度から文書の特徴を判断することができるようになる. それにより文書における出現回数を特徴量とすることで文章の感情に強弱をつけることができると考えられる.

WordNet-affect とは単語に感情やしぐさ，情動をタグ付けした語彙データベースであり，菅原 [7] はこれを用いて品詞に感情尺度を付与した辞書,「感情語辞書」の生成を実現した．感情語辞書は分類された感情と感情の強度を 0.0〜1.0 で表したものを併せた分類結果を出力としており，これを用いることで本研究においても感情の強弱の付与が可能となると考える.

8 おわりに

文章をフレーズ単位に分割しフレーズ内容に合致した感情音声で読み上げるシステムの提案について論じてきた. 図 1 より, 本研究では感情分類器にナイーブベイズ分類, 音声出力システムに HMM 音声合成, データセットは様々な分野の文を抽出し, コーパス構築を行っている. 今後はデータセットや分類・出力システムを置き換えることによる利点などを研究していき, 様々なテキストデータにおいて適切なサブシステムを切り替えられるようなシステムを考案していきたい.

参考文献

[1] 山岸順一，大西浩二，益子貴史，小林隆夫：HMM音声合成における発話スタイルの制御，情報処理学会第 65 回全国大会

[2] 山本麻由，土屋誠司，黒岩眞吾，任福継：感情コーパス構築のための文中の語に基づく感情分類手法，社会法人情報処理学会

[3] コーパス検索アプリケーション「中納言」, 国立国語研究所：https://chunagon.ninjal.ac.jp

[4] 山本 和英，SNOW D18 日本語感情表現辞書，長岡技術科学大学 電気電子情報工学専攻，2018.7.24

[5] 徳田恵一, 大浦圭一郎, 橋本佳, 南角吉彦:隠れマルコフモデルに基づく日本語音声合成ソフトウェア入門

[6] 山本和英:日本語処理を用意にする活用形態素の提案 長岡技術科学大学 Japio2016

[7] 菅原久嗣，感情語辞書を用いた日本語テキストからの感情抽出，東京大学大学院 情報理工学系研究科 電子情報学専攻