

8M-1 形態素情報を用いたテキストからの感情情報の抽出*

長谷川 隆明[†] 小原 永
NTT サイバースペース研究所[‡]
E-mail: hasegawa@isl.ntt.co.jp

1 はじめに

近年、メールやホームページの読み上げを行うサービスが多数社会に登場しているが、任意のテキストに対して喜怒哀楽が表現できるほど豊かな抑揚の付いた合成音声を実現するレベルには至っていない。これを実現するには、テキスト解析時に意図を表現するモダリティや感情という情報を抽出しておくことが不可欠である。しかしながら、従来のモダリティの分析や抽出に関する研究 [3, 4] で報告されているような、助動詞などの特定の品詞や文末表現および述語のみに着目した方法では不十分であり、これらに出現するモダリティに限定されることなく、これら以外に出現する多様な感情情報を抽出しなければならない。そこで、本稿では、形態素情報を用いることにより、モダリティを含めた感情情報をテキストから抽出する手法を提案する。

2 感情情報の抽出

テキストから感情情報を抽出するための手法の概要を図 1 に示す。本稿で提案する手法は、初期化時における感情情報の定義と感情情報パターンへのコンパイルおよび実行時における形態素情報によるマッチングの各処理からなる。

2.1 感情情報の定義

初期化時には、はじめに、感情情報の定義として、各感情情報の種類とそれに対応する表現をテキストレベルで記述する。記述可能な表現は、形態素の表記、品詞、読み、一般名詞カテゴリ¹の属性を用いた表現である。感情情報の定義の具体例を図 2 に示す。

本手法における感情情報の定義に関する記述の特長は、(1) 属性に表記を用いる場合、複数の形態素からな

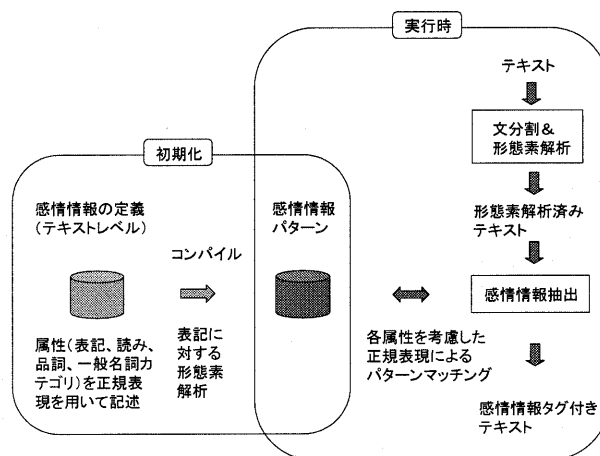


図 1: 感情情報抽出の手法の概要

る表記の単位で記述が可能、(2) 形態素の属性の並びによる組み合わせによる記述が可能、(3) 論理和 ($\{a|b\}$)、先頭マッチ (^)、末尾マッチ (\$)、否定 (!) 等の正規表現を用いた記述が可能、の 3 点である。つまり、ある表記を定義するときあらかじめ形態素の分割のされ方がわからなくてもよく、品詞や一般名詞カテゴリの情報を組み合わせることができ、形態素の位置等の特定も可能である。

2.2 感情情報パターンへのコンパイル

実行時に行われる入力されたテキストとのパターンマッチングの準備として、テキストレベルで記述された感情情報の定義のコンパイルによって感情情報パターンを生成する。パターンの生成の順序として、まず論理和などの正規表現が存在すればこれを展開する。次に、感情情報の各定義の表記に対して正規表現の記号を除いた文字列に対して JTAG[1] による形態素解析を行い、正規表現の記号の意味を考慮した、各形態素の表記からなるパターンに分割する。例えば、図 2 の [怒り] の記述例では、図 3 に示すようにパターンが生成される。

*Emotions Extraction from Text Using Morphological Information

[†]Takaaki Hasegawa and Hisashi Ohara

[‡]Cyber Space Laboratories, NTT

¹ 本稿における一般名詞カテゴリの体系は、NTT シソーラス [2] の一般名詞意味属性の体系に準じている。

[喜び]	表記 = { 美味しい おいしい 美味しかった おいしかった }
[喜び]	カテゴリ = { 喜び 成功 幸運 }
[哀しみ]	品詞 = 名詞:動作, 表記 = して { くれず くれません もらえず もらえません }
[哀しみ]	カテゴリ = { 問題 悪意 事故 }, 表記 = { が は 等 など の }, 表記 = { 無い 無かった }
[怒り]	表記 = いい加減にし { て なさい ろ }
[驚き]	表記 = { えっ えーっ エッ エーッ } { \$ 、 。 }

図 2: 感情情報の定義例

[感情情報の定義]
 表記=いい加減にし{て|なさい|ろ}
 ↓(論理和の展開)
 表記={いい加減にして|いい加減にしなさい|いい加減にしろ}
 ↓(形態素解析)
 [感情情報パターン]
 {表記=いい加減, 表記=に, 表記=して|表記=いい加減, 表記=に, 表記=し, 表記=なさい|表記=いい加減, 表記=に, 表記=しろ}

図 3: パターン生成の例

2.3 形態素情報によるマッチング

実行時では、はじめに、入力されたテキストに対して一文等を一行とする文分割処理の後に、形態素解析によりテキストを形態素に分割し、各形態素の表記、品詞、読み、一般名詞カテゴリの属性情報を得る。次に、入力テキストの形態素と感情情報パターンを用いて、形態素の属性情報を用いたパターンマッチングを行う。このとき、感情情報パターンで指定されている表記、品詞、読み、一般名詞カテゴリといった形態素の各属性情報が形態素の属性情報を伴った入力テキストの形態素に一致するかどうかの判定を、感情情報パターンに存在するあるパターンのすべての形態素の属性情報の並びに対して行い、各々がすべて入力テキストの形態素の並びに一致する場合にそのパターンがマッチしたとする。

感情情報パターンに正規表現が存在する場合は、次のようにする。パターンの1つに論理和が用いられている場合は、論理和を構成するいずれかのパターンの形態素の並びがすべてマッチするときに全体の論理和のパターンがマッチしたと見なす。さらに、先頭マッチ、末尾マッチ、否定については、先頭マッチの記号が存在すれば文の先頭からパターンマッチングを行い、末尾マッチの記号が存在すれば文末とのパターンマッチングを行う。否定の記号が存在する場合は、否定の記号が付けられている属性情報以外であればマッチしたとする。

3 評価

提案した手法を評価するために、電子メールのテキストを対象とした評価実験を行った。評価実験のため用意した感情情報の種類は、[喜び]、[怒り]、[哀しみ]、[楽しみ]、[驚き]の5つである。感情情報の定義をコンパイルした後の感情情報パターンの数は、図2に示した定義例を含めて、計828個である。これらは論理和が用いられているパターンの数については、すべて展開したパターンの数として計算している。実験に用いた電子メールの数は、モダリティ情報の定義を記述するために用いたクロードデータが77通、用いなかったオープンデータが63通である。これらはビジネスのメール、プライベートのメール、商用サービスのメーリングリストからの配信等が混在したものである。

実験の結果から、感情情報のすべてを対象としたときのF-measureの値がクロードデータで87.8、オープンデータで79.0という高い値が得られた。

4 おわりに

本稿では、形態素の持つ属性情報を用いることによって、テキストから感情情報を抽出する手法を提案した。電子メールのテキストを対象とした実験の結果、高い値のF-measureが得られたことから、本稿で提案した手法は良好な精度が得られることを検証した。現在、音韻毎のピッチ周波数や振幅値(パワー)などの韻律特性を韻律制御コマンドとして記述できる言語であるMSCL[5]との結合により、入力されたテキストから感情情報に応じた抑揚付きの合成音声出力するソフトウェアの開発を進めている。このソフトウェアによって、任意のテキストから適切な抑揚付きの合成音声生成できるので、様々な応用に利用することが可能である。

参考文献

- [1] Fuchi, T. and Takagi, S.: Japanese Morphological Analyzer Using Word Co-occurrence, *Proc. COLING-ACL'98*, pp.409-413, 1998.
- [2] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林: 日本語語彙体系 1 意味体系, 岩波書店, 1997.
- [3] 乾, 内元, 井佐原: モダリティ分析に基づく自由回答アンケートの分類, 言語処理学会第4回年次大会発表論文集, pp.540-543, 1998.
- [4] 諸橋, 那須川: テキストマイニング: 膨大な文書データからの知識獲得—意図の認識—, 情報処理学会第57回全国大会 5K-03, 1998.
- [5] 水野, 中野: 合成音声制御のための階層型記述言語 MSCL, 人工知能学会研究会資料 SIG-SLUD-9703, pp.7-12, 1998.