

## N-gram 出現頻度を用いた感情類似度計算

三品 賢一<sup>†</sup> 土屋 誠司<sup>††</sup> 黒岩 眞吾<sup>††</sup> 任 福継<sup>††</sup>

<sup>†</sup> 徳島大学大学院先端技術教育学部

<sup>††</sup> 徳島大学大学院ソシオテクノサイエンス研究部

770-8506 徳島県徳島市南常三島 2-1

E-mail: †mishina@is.tokushima-u.ac.jp, ††{tsuchiya,kuroiwa,ren}@is.tokushima-u.ac.jp

あらまし 従来、文が表現する感情を推定する手法では、推定できる感情の種類がわずかであったり、一つ一つの単語ごとに感情別の重みを付与した辞書を構築し、推定に利用するものが多かった。そこで我々は、推定できる感情の種類を容易に増やすことができ、また感情別の重みの付与を単語 N-gram で行うことで文が指す内容に対する話し手の判断や心的態度を表すモダリティなどの感情表現も利用する、従来手法よりも高い精度で推定可能な感情推定手法の提案を目指している。このような推定手法を実現するため、我々は2つの文が表現している感情がどれほど類似しているかを計算する感情類似度計算手法を過去に提案した。感情類似度は、あらかじめ用意した、文を感情別に分類した複数のコーパス(感情コーパス)を用い、機械翻訳システムの翻訳精度を求めめる尺度である BLEU を基に、入力文と感情別に分類された文との類似度を計算することで求める。本稿で我々は、従来の BLEU を用いる感情推定よりも高い精度で推定を行うために、感情コーパス別に単語 N-gram の出現頻度を求めた辞書を従来手法に導入した新たな手法を提案する。提案手法の性能を調べるため、入力文から感情類似度を求め、最も感情類似度が高くなった感情と、人手で判断した入力文の感情の一致率を求めめる実験を行った。その結果、従来の BLEU による類似度計算を用いた手法に比べ、提案手法では 20.59%一致率が向上した。

キーワード 感性情報処理, 類似度計算, 用例ベース, N-gram

## An Emotion Similarity Calculation Using N-gram Frequency

Kenichi MISHINA<sup>†</sup>, Seiji TSUCHIYA<sup>††</sup>, Shingo KUROIWA<sup>††</sup>, and Fuji REN<sup>††</sup>

<sup>†</sup> Graduate School of Advanced Technology and Science, The University of Tokushima

<sup>††</sup> Institute of Technology and Science

2-1 Minamijosanjima, Tokushima 770-8506

E-mail: †mishina@is.tokushima-u.ac.jp, ††{tsuchiya,kuroiwa,ren}@is.tokushima-u.ac.jp

**Abstract** Existing methods to estimate emotions of a sentence can estimate a few kinds of emotion, and many methods use a dictionary that emotion weights is related every words. Our aim is to propose a method which can estimate with high precision using feeling expression (modality et al.) and can add new estimatable emotion easily. we gave emotion weights to N-gram to use feeling expression. To propose such method, we proposed an emotion similarity calculation. This method calculates similarity between represented emotions from two sentences. In the emotion similarity calculation, we use similarity between an input sentence and classified sentence by emotions. Classified sentences are in emotion corpora. The calculation formula is based on BLEU. BLEU is machine translation evaluation method. In this paper, we propose a new emotion similarity calculation method. This method can estimate with high precision compared to the past one. This method uses N-gram frequency dictionaries made from each emotion corpora. To examine the precision of the method, we evaluated ratio that emotions of highest emotion similarities corresponds emotions of the input. Emotions of the input was decided by human. As the results, the ratio was improved 20.59.

**Key words** affective computing, similar calculation, example based, N-gram

## 1. まえがき

近年、人の感性に刺激を与えることで感情の変化を促すエージェントシステムなどを開発する需要が高まっている [1]。感情の変化を促すことで、例えばドライバーの気分を落ち着かせたりするなど、様々な仕事やオペレーションに適した状態へ感情を変化させることが出来る。このようなエージェントシステムを構築するためには、人がどういった感情のときにどういった言動を取るのかを予測する必要がある。この予測に基づきエージェントシステムは人の感情状態を知り、人の感情に適した状態へ変化させるための行動に移すことができる。

このようなエージェントシステムを実現するため、我々は会話から話し手の感情を推定するための技術の実現を目指している。会話から感情を推定するためには文が表現する感情を推定する技術が必要と考え、現在我々は文から感情を推定する技術を研究している [2], [3]。またこれまでも文から感情を推定する技術に関する研究は行われている。

目良ら [4] は、複数の事象の格フレームタイプのうち、どれに入力文が当てはまるかを判定し、該当した格フレームタイプに対応する情緒計算式を用いて発話者の感情が快か不快かを判定する手法を提案している。この手法では、あらかじめ用意した情緒感式の外に、ユーザの嗜好情報を基にした、単語に対する好感度データを用いる。好感度を与える単語は、格フレームタイプを構成するのに必要な語である名詞や動詞、形容詞、副詞などの内容語となる。

Jianhua [5] は感情を表現すると考えられる単語に 1 つ以上の感情別重みを付与しておき、ある文において発話者が表現している感情を推定する “Emotion eStimation Net (ESiN)” を提案している。この手法では、感情や感情の強度を表現する単語の辞書を複数人で構築して感情推定に用いており、推定できる感情の種類は joy, sadness, anger, surprise, hate, fear の 6 種類である。単語に付与する感情別重みの種類を変えれば、推定する感情の種類を変えることができると考えられる。構築している単語の辞書には、感情を表現している語 (“unhappy” など)、程度を表す語 (“very” や “so”), 感情を表す行動や状態を示す語 (“hate” や “asperity”) や、これら以外の内容語を登録する。

高村ら [6] は文がポジティブとネガティブのいずれを示しているかを推定するためのリソースを半自動で構築する手法を提案している。この手法では、少量の単語に感情極性 (ポジティブかネガティブか) を付与しておき、自動で大量の語に感情極性を付与する。数語から十数語という少数の単語を種として、高い正解率での単語の感情極性判定を実現している。この手法で感情極性を付与する単語は内容語となっている。

一方で本論文で提案する手法は、文の類似度を計算する手法を取り入れている。文の類似度を計算する手法として、下畑ら [7] が共通 N-gram を用いた類似度計算を行っている。これは機械翻訳文の自動評価方法である BLEU [8] と同じ式を用いている。さらに、単語の品詞ごとに重み付けを行い、類似度計算の精度を向上させている。共通 N-gram を用いていること

で、共通単語や語順も考慮できる。しかし、入力文の単語数が少ない場合、表現する感情が全く異なっているにもかかわらず、同等の類似度を得るケースが存在する。例えば、入力文として “愚かだ” という文が与えられたとき、入力文と “好きだ” との類似度は、入力文と “嫌いだ” との類似度と同じになる。これは 3 文とも “だ” を含んでいるが、文頭の単語が異なっているため、“好きだ” との類似度と、入力文と “嫌いだ” との類似度が同じになる。表現する感情も考慮に入れる場合、これらの類似度が同じになるべきではない。

過去に我々は、BLEU を文と文の類似度を計算するための式として用いた感情推定手法 [9] を提案している。この手法では、感情ごとに文を分類したコーパス (感情コーパス) を用い、入力文と最も類似度の高い文が含まれるコーパスの種類から感情を推定する。そのためコーパスの分類数が推定可能な感情の種類の数となり、複数種類の感情が推定可能となる。また感情推定の際に内容語に加え、助詞や助動詞などの付属語もすべて利用しているため、文が指す内容に対する話し手の判断や心的態度を表すモダリティなどに含まれる感情表現も考慮でき、感情推定精度の向上が期待できる。この手法で計算する類似度を “感情類似度” と呼ぶ。図 1 に感情類似度計算を用いた感情推定の流れを示す。例えば、“怒り” の文を集めて構築したコーパスの中に、入力文と最も類似した文があるならば、入力文が表現する感情は “怒り” となる。

本論文では、感情コーパス別に単語 N-gram の出現頻度を求めた辞書 (N-gram 出現頻度辞書) を導入した BLEU を、従来の感情類似度計算に用いた手法を提案する。この辞書を導入することで、従来手法に比べ推定精度が向上することを示す。まず 2 章で感情類似度計算に用いた感情コーパスと、N-gram 出現頻度辞書、感情類似度計算式について述べる。そして 3 章で提案手法による推定成功率の評価結果を述べる。

## 2. 感情類似度計算

### 2.1 感情コーパス

提案手法で用いるコーパスは、Web 上から収集した文から成る。収集者は文の発話者が表現していると判定した感情ごとに分類した。分類する感情の種類は収集者が決めた。判定した感情の種類が複数の場合を考慮して、分類先は 2 種類以上も許容した。また、ひとつの文の感情の判定は収集者一名が行った。そのため複数の収集者がひとつの文の感情を判定するといった方法はとっていない。現在行っている分類のすべての種類と文数は表 1 の通りであり、合計 8 名の収集者で作業を行った。また、収集元は主に Yahoo! の掲示板 (<http://messages.yahoo.co.jp>) である。

### 2.2 N-gram 出現頻度辞書

N-gram 出現頻度辞書とは、各感情コーパスにおける N-gram の出現頻度を登録した辞書である。表 2 に各感情コーパスにおける N-gram の出現頻度の一部を示す。例えば、副詞である “すっきり” は、喜びのコーパスにおける出現頻度が 3 となり、怒り、嫌悪、希望の感情コーパスにおける出現頻度は 0 となった。

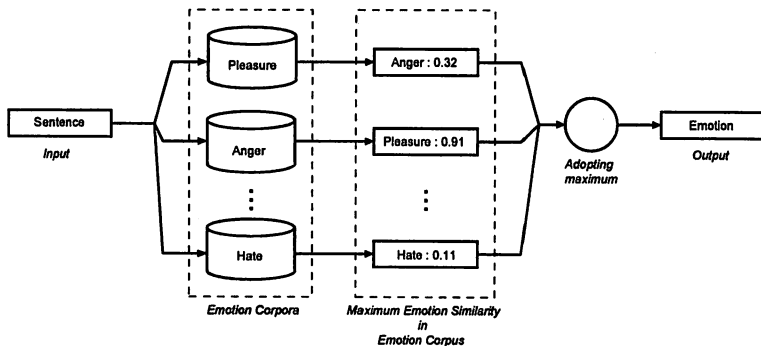


図1 感情類似度を用いた感情推定

Fig.1 An emotion estimation using emotion similarity

表2 各感情コーパスにおける N-gram 出現頻度の一部

Table 2 Part of N-gram frequency on the emotion corpora

N-gram	コーパス別 N-gram 出現数			
	喜び	怒り	嫉妬	希望
<副詞>すっきり</副詞>	3	0	0	0
<名詞>一緒</名詞>	2	6	8	15
<名詞>ためえ</名詞>	0	4	1	0
<副詞>何故</副詞>	0	5	0	0
<助動詞>たい</助動詞>, <助詞>なあ</助詞>, <記号>。</記号>	0	0	0	8
<助動詞>たい</助動詞>, <助詞>と</助詞>, <助詞>思い</助詞>	0	0	0	4
<名詞>不愉快</名詞>, <助動詞>です</助動詞>, <助詞>ね</助詞>, <記号>。</記号>	0	0	2	0
<助詞>は</助詞>, <助詞>は</助詞>, <助詞>は</助詞>, <助詞>は</助詞>	5	0	0	0
<助詞>は</助詞>, <助詞>は</助詞>, <助詞>は</助詞>, <助詞>は</助詞>, <助詞>は</助詞>	3	0	0	0

表1 感情コーパスの文数

Table 1 Count of sentences in corpora that was classified in emotions

怒り	2394	喜び	1354	嫉妬	1173
平静	1092	希望	838	不安	620
侮蔑	332	驚き	331	感嘆	330
悲しみ	269	申し訳なさ	253	疑問	247
落胆	241	共感	207	卑下	199
苦痛	199	困惑	138	呆れ	129
羨み	93	悔しさ	86	焦燥	75
安堵	65	恐怖	64	恥	57
励まし	48	不満	46	哀れみ	23
奮起	22	孤独	18	同情	17
深刻	17	迷い	14	諦め	14
企み	11	信頼	3	誇り	3
嫉妬	1	疲労	1		

### 2.3 BLEU

感情類似度計算式は、BLEU をベースにした。BLEU は機械翻訳システムが出力した複数の翻訳候補文から、システムの翻訳精度として評価するための尺度である。提案手法では BLEU を、文同士の類似度として扱う。BLEU を用いた理由は次の 2 点である。まず BLEU は翻訳精度を人間に近い評価スコアを出せるという特徴があるということ、そして BLEU を用いた

類似度計算は熟語だけでなく文末表現も考慮できると考えたからである。

BLEU は次のように定義されている [8]。

$$BLEU = BP \cdot \exp \left( \sum_{n=1}^N w_n \log p_n \right) \quad (1)$$

ここで  $w_n$  は  $1/N$  となる。[8] では、 $N = 4$  を基本的に用いている。 $p_n$  は機械翻訳文と人による翻訳文における共通 N-gram 数の適合率 (N-gram 適合率) であり、BP は機械翻訳文が人による翻訳文に比べて簡潔すぎることによる適合率のペナルティとなる。以下、N-gram 適合率と、適合率のペナルティについて述べる。

#### 2.3.1 N-gram 適合率

N-gram 適合率  $p_n$  は、機械翻訳文と人による翻訳文が共通な N-gram をどれほど持っているかを表す値である。共通な N-gram が多いほど  $p_n$  は大きくなる。

N-gram 適合率の計算には、機械翻訳システムが出力した 1 つ以上の翻訳文 (Candidate) と、人手により用意した 1 つ以上の翻訳文 (Reference) を用いる。ここでは例として、次のような Candidate と Reference があるとする。

Candidate1 : That dog is on this bed.

Candidate2 : This is a cat.

表 3 Candidate1 における各 1-gram の  $Count, Count_{clip}$   
Table 3  $Count, Count_{clip}$  of 1-gram in Candidate1

	Count, Count <sub>clip</sub> of 1-gram in Candidate1					
	that	dog	is	on	this	bed
Count <sub>clip</sub>	1	1	1	1	0	1
Count	1	1	1	1	1	1

表 4 Candidate2 における各 1-gram の  $Count, Count_{clip}$   
Table 4  $Count, Count_{clip}$  of 1-gram in Candidate2

	Count, Count <sub>clip</sub> of 1-gram in Candidate2			
	this	is	a	cat
Count <sub>clip</sub>	0	1	0	0
Count	1	1	1	1

表 5 Candidate1 における各 2-gram の  $Count, Count_{clip}$   
Table 5  $Count, Count_{clip}$  of 2-gram in Candidate1

	Count, Count <sub>clip</sub> of 2-gram in Candidate1					
	that dog	dog is	is on	on this	this bed	
Count <sub>clip</sub>	1	1	1	0	0	
Count	1	1	1	1	1	

表 6 Candidate2 における各 2-gram の  $Count, Count_{clip}$   
Table 6  $Count, Count_{clip}$  of 2-gram in Candidate2

	Count, Count <sub>clip</sub> of 2-gram in Candidate2		
	this is	is a	a cat
Count <sub>clip</sub>	0	0	0
Count	1	1	1

Reference1 : The dog is on the bed.

Reference2 : There is a dog on the bed.

これらの Candidate, Reference を用いて,  $p_n$  は次のように定義されている。

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-gram} \in C} Count_{clip}(n\text{-gram})}{\sum_{C \in \{Candidates\}} \sum_{n\text{-gram} \in C} Count(n\text{-gram})} \quad (2)$$

ここで  $\{Candidates\}$  を Candidate の集合 (Candidate1, Candidate2 を要素とする集合),  $C$  を  $\{Candidates\}$  の要素 (Candidate1 や Candidate2),  $n\text{-gram}$  を  $C$  に含まれる連続する  $n$  個の形態素から作られる文字列とする。Count 関数は, 1つの Candidate における  $n\text{-gram}$  の出現数を返し,  $Count_{clip}$  関数は  $\min(Count, Max\_Ref\_Count)$  を返す。Max\_Ref\_Count 関数はすべての Reference ごとにある  $n\text{-gram}$  の出現数を求め, その中で最も多い出現数を返す。機械翻訳文 Candidate が, 人が翻訳した文 Reference と共通な N-gram を持っていなければ  $Count_{clip}$  は 0 になるため,  $p_n$  は機械翻訳文と人の翻訳文がどれほど共通な N-gram を持っているかの指標となる。

N-gram 適合率  $p_n$  を求める例として, 機械翻訳システムの出力した翻訳文から Unigram の適合率  $p_1$  と Bigram の適合

率  $p_2$  を計算する。表 3 に Candidate1 に対する Unigram の  $Count$  と  $Count_{clip}$  の値を, 表 4 に Candidate2 に対する Unigram の  $Count$  と  $Count_{clip}$  の値を示す。表 3, 4 と式 (2) から, Unigram 適合率  $p_1$  は  $3/5$  であるということがわかる。

また表 5 に Candidate1 に対する Bigram の  $Count$  と  $Count_{clip}$  の値を, 表 6 に Candidate2 に対する Bigram の  $Count$  と  $Count_{clip}$  の値を示す。表 5, 6 と式 (2) から, Bigram 適合率  $p_2$  は  $3/8$  であるということがわかる。

### 2.3.2 適合率のペナルティ

ここでは機械翻訳文 Candidate が人の翻訳文 Reference に比べて簡潔すぎることによる適合率のペナルティ (brevity penalty, 以下 BP) について説明する。BP を計算するには, まず形態素数が最も近い Reference と Candidate のペアを探す。例えば, Candidate1 の形態素数が最も近いのは Reference1 であり, Candidate2 の形態素数が最も近いのも Reference1 である。このようなペアを探し出し, 各ペアに含まれるすべての Reference の形態素数の総和を  $r$ , Candidate の形態素数の総和を  $c$  とし, BP は次のように定義される。

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

Candidate1, 2 と Reference1, 2 から BP を求めた場合,  $c = 6 + 4$  となり,  $r = 6 + 6$  となるので,  $BP = e^{1-12/10} < 1$  となる。これは, Candidate2 が短すぎるため, 適合率ペナルティが課せられることを意味する。

もし形態素数が 3 と 5 の Reference が存在していた場合, いずれの文も Candidate2 の形態素数に最も近くなるが, [8] では BP が減衰するように Reference を選ぶとしているので, 形態素数が 5 の Reference を Candidate2 のペアの相手として選ぶことになる。

### 2.4 感情類似度計算式

本論文で提案する感情類似度は, 2つの文が表現している感情がどれほど類似しているかを表す値である。提案手法では感情類似度を用いることで, 例えば入力文が怒りを表現した文とどの程度類似しているかを求める。

感情類似度計算式において, N-gram 同士の区別を行う際に用いた情報は, 形態素文字列及び形態素の品詞である。N-gram が文頭もしくは文末に出現したか, といった位置情報は用いていない。

感情類似度計算式に N-gram 出現頻度を用いた理由は, N-gram の出現頻度が感情コーパスごとに異なれば, 頻度情報を感情別の重みとして扱うことができ, 感情を考慮した類似度計算ができると考えたためである。例えば, “すっきり” という形容詞は, 嬉しいときや喜んでいときに使われやすく, 怒っているときにはあまり使われないと考えられる。実際に, “喜び”, “怒り”, “嫌悪”, “希望” の 4 種類のコーパスの中から, 各ランダムに 838 文選択し, N-gram 出現頻度を調べてみた。表 2 に出現頻度データの一部を示す。なお, 形態素の前後に品詞のタグを付けている。この表より, “すっきり” は, 喜び

のコーパスでの出現頻度が高いことがわかる。

感情類似度計算式は BLEU がベースになっている。式 (1) をもとに、感情類似度 ES を次のとおりに定義する。

$$\begin{aligned} es &= BP \cdot \exp \left( \sum_{n=1}^N w_n \log p_n \right) \\ &= BP \cdot \exp \left( \sum_{n=1}^N \frac{1}{N} (1 - e\bar{w}) \log p_n \right) \\ ES &= \max_{Candidate \in E} es \end{aligned} \quad (3)$$

ここで  $E$  を、求める感情類似度 ES の感情に対応する感情コーパス (以後 TargetCorpus と呼ぶ)、 $Candidate$  を  $E$  のうちの一文とする。ES への入力となる感情推定対象文は、 $e\bar{w}$  と  $p_n$  の計算に用いる。また  $e\bar{w}$  は次の通りに定義した。

$$\begin{aligned} ew &= \frac{nf_E(n\text{-gram})}{\sum_{E' \in \{Corpora\}} nf_{E'}(n\text{-gram})} \\ e\bar{w} &= \frac{1}{N\text{gramCount}} \sum_{n\text{-gram} \in Input} ew(n\text{-gram}) \end{aligned} \quad (4)$$

ここで  $Input$  を感情推定対象文、 $N\text{gramCount}$  を  $Input$  から得られる  $N$ -gram の個数、 $\{Corpora\}$  を TargetCorpus を含むすべての感情コーパス、 $E$  を TargetCorpus、 $nf_X$  を感情コーパス  $X$  における  $n$ -gram の出現頻度を返す関数とする。 $nf_X$  は  $N$ -gram 出現頻度辞書を用いて  $n$ -gram の出現頻度を返す。

$ew$  はすべての感情コーパスのうち TargetCorpus に出現する  $n$ -gram の出現率を返す。 $e\bar{w}$  は、感情推定対象文  $Input$  の  $N$ -gram が TargetCorpus に多く出現していれば大きくなる。 $e\bar{w}$  が大きくなると、ES も大きくなるので、 $e\bar{w}$  は感情推定対象文  $Input$  の感情と TargetCorpus に対応する感情が一致すれば感情類似度 ES を大きくするといった特徴を持つ式となる。

また感情類似度計算式では  $p_n$  を求める計算式、式 (2) における  $\{Candidates\}$  を  $Candidate$  のみとした。そのため  $p_n$  は次のように簡略化される。

$$\begin{aligned} p_n &= \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-gram} \in C} Count_{clip}(n\text{-gram})}{\sum_{C \in \{Candidates\}} \sum_{n\text{-gram} \in C} Count(n\text{-gram})} \\ &= \frac{\sum_{n\text{-gram} \in Candidate} Count_{clip}(n\text{-gram})}{\sum_{n\text{-gram} \in Candidate} Count(n\text{-gram})} \end{aligned} \quad (5)$$

ただし、 $Count_{clip}$  関数は  $\min(Count, Max\_Input\_Count)$  を返すとし、 $Max\_Input\_Count$  関数は感情推定対象文にある  $n$ -gram の出現数を求め、その中で最も多い出現数を返すとする。

### 3. 評価実験

#### 3.1 提案手法と BLEU の比較

提案手法が従来手法と比べ、どれほど感情推定性能を向上させることができるかを調べるための評価実験を行った。評価方法として、leave-one-out cross validation 法を用いた。実験の手順は以下の通りである。

[step 1] 感情ごとに分類した、いくつかのコーパスを用意する。  
[step 2] 用意したコーパスのうち一つから、文を一つ取り出し、この文を入力とする。

[step 3] 各コーパスごとの  $N$ -gram 出現頻度を計算する。

[step 4] 各コーパスごとに最も大きい入力文との感情類似度を求める。

[step 5] 感情類似度が高かった順番でコーパスに対応する感情を出力する。

step 2 から 5 を、毎回入力文を変えて繰り返し行う。感情コーパス構築時に、文を複数種類への分類を認めているので、入力文の感情は 1 つ以上となる。そのため、入力文の感情が出力感情の一位から連続で出現すれば推定成功とした。例えば、入力文の感情が“喜び”と“希望”であり、出力感情の一位と二位が“希望”と“喜び”であった場合は成功とした。出力感情の一位が“希望”、二位が“怒り”、三位が“喜び”の場合は失敗とした。また“希望”と“喜び”、“怒り”がいずれも一位であった場合などは、入力文の感情と同じ順位に、入力文の感情とは異なる感情が出現していることから失敗とした。

この実験により、提案手法を用いた感情推定成功率を求めた。また、step 4 の感情類似度を求める箇所を BLEU を求めるように変更した同様の実験も行い、BLEU を用いた感情推定成功率も求めた。また求めた感情類似度と BLEU は、 $N$  が 1 から 6 までのものである。本論文では  $N = 1$  とした感情類似度は ES1、BLEU は BLEU1 と表記する。なお、感情類似度を求める際、step 3 で計算した  $N$ -gram 出現頻度から入力文における  $N$ -gram 出現頻度情報を差し引いた。

実験に用いたコーパスは、収集数が多い喜び、怒り、嫌悪、希望で、各 838 文である。平静を除いた理由は、平静の文には感情をあまり表現していない文が多いので、平静の文と平静以外の文との感情類似度と、平静の文同士の感情類似度に大きな違いが見られず、その結果 BLEU と ES の比較が行うことができないと考えたためである。

##### 3.1.1 比較結果

表 7 に推定成功率を示す。表 7 より、BLEU による推定成功率の中で最も結果がよかった BLEU2 に比べ、提案手法による推定成功率の中で最も結果がよかった ES2 では推定成功率が 20.59% 向上していることがわかる。 $N$ -gram 出現頻度が感情別の重みとして働き、推定成功率を向上させることができたと考えられる。

また BLEU、ES とともに  $N = 2$  のときが最も推定成功率が高く、その後  $N$  が大きくなるにつれ推定成功率は減少している。ES の推定成功率が減少した理由として、次の二つが考えられる。

表 7 感情推定成功率

Table 7 Emotion estimation success ratios

手法	推定成功率 [%]
BLEU1	50.65
BLEU2	55.61
BLEU3	54.68
BLEU4	50.53
BLEU5	44.77
BLEU6	41.61
ES1	74.62
ES2	76.20
ES3	72.02
ES4	70.11
ES5	68.50
ES6	67.67

一つ目は、入力文との感情類似度が最も大きくなったコーパス中の文を確認したところ、Nが大きいくときほど感情類似度が最も大きくなるコーパス中の文の長さが短くなる傾向が BLEU, ES ともに見られた。感情類似度が最も大きくなるコーパス中の文の長さが短くなると、感情類似度が最も大きくなるコーパス中の文同士が持つ N-gram の違いが減り、各感情の感情類似度の差が小さくなる。そのため入力文の感情と異なる感情が出力感情の上位に表れやすくなり、その結果 N が大きいほど推定成功率が減少したと考えられる。BLEU と同様に、ES も N が 2 より大きくなるにつれて推定成功率が減少しているの、N が大きいときに推定成功率を下げる要因が BLEU の式自体にあると考えられる。

二つ目は、N が大きいとき、例えば 6-gram などはコーパス中での出現頻度は 0 になることが多くなり、その結果、ES での N-gram 適合率に対する重みが BLEU の重みと等しくなり、BLEU と同様に N が大きくなるにつれて ES も推定成功率が減少したと考えられる。この問題を解決するために、スムージング [10] を提案手法に導入することが考えられる。N-gram モデルを構築する際に、学習データに存在しない N-gram の出現確率を求める問題としてゼロ頻度問題が知られており、この問題を解決するためにスムージングと呼ばれる手法が提案されている。このスムージングを提案手法に導入することで、ES での N-gram 適合率に対する重みが BLEU の重みと等しくなることが回避でき、N が大きいときの推定成功率を改善できると考えられる。

#### 4. むすび

今回、2つの文が表現している感情がどれほど類似しているかを計算する新たな感情類似度計算手法として、N-gram 出現頻度辞書を BLEU に導入した手法を提案した。実験の結果、従来手法に比べ、推定精度が 20.59%向上したことを確認した。

今後の課題として、感情コーパスの拡充、BLEU 以外の類似度計算 [11], [12], [13] を用いた感情類似度計算式による推定精度評価、スムージングを用いた N-gram 出現頻度の利用があげられる。また提案手法では入力された一つの文をもとに感情

を推定しているが、現実には現在の感情状態が過去の感情状態に影響を受けることも考えられる。例えば、過去の感情状態が“悲しい”状態であったとき、“楽しい”感情を表現する言葉を口にした場合、発話者の感情は本当に“楽しい”わけではなく、単なる“空元気”であることも考えられる。従って、会話の流れや過去の感情状態の情報も用いた感情推定も出来るよう、提案手法や提案手法を用いた感情推定手法を拡張していきたい。

#### 文 献

- [1] C.M. Jones and I.M. Jonsson, “Detecting emotions in conversations between driver and in-car information systems,” *Affective Computing Intelligent Interaction 2005*, pp.22–24, October 2005.
- [2] K. Matsumoto, J. Minato, F. Ren, and S. Kuroiwa, “Estimating human emotions using wording and sentence patterns,” *IEEE International Conference on Information Acquisition*, vol.2005, no.1, pp.421–426, June 2005.
- [3] H. Xiang, P. Jiang, S. Xiao, F. Ren, and S. Kuroiwa, “An emotion information processing model based on a mental state transition network,” *IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE'05)*, pp.668–673, Oct 2005.
- [4] K. Mera, T. Ichimura, T. Aizawa, and T. Yamashita, “Invoking emotions in a dialog system based on word-impressions,” *Journal of Japan Society of Artificial Intelligence*, vol.17, no.3, pp.186–195, 2002.
- [5] J. Tao, “Context based emotion detection from text input,” *INTERSPEECH 2004 - ICSLP*, pp.1337–1340, 8th International Conference on Spoken Language Processing, Oct 2004.
- [6] H. Takamura, T. Inui, and M. Okumura, “Extracting semantic orientation of words using spin model,” *IPSJ Journal*, vol.47, no.02, pp.627–637, 2006.
- [7] M. Shimohata, E. Sumita, and Y. Matsumoto, “A method for retrieving a similar sentence and its application to speech translation,” *Journal of Natural Language Processing*, vol.11, no.4, pp.105–126, 2004.
- [8] K. Papineni, S. Roukos, T. Ward, and W.J. Zhu, “Bleu: a method for automatic evaluation of machine translation,” *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp.311–318, Association for Computational Linguistics, 2001.
- [9] K. Mishina, F. Ren, and S. Kuroiwa, “An emotion similarity calculation using dialog sentence corpora,” *The International Symposium on Artificial Intelligence and Affective Computing (AIAC-06)*, 2006.
- [10] K. Kita, *Probabilistic Language Model*, University of Tokyo Press, 1999.
- [11] C.Y. Lin and E. Hovy, “Automatic evaluation of summaries using n-gram co-occurrence statistics,” *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, Morristown, NJ, USA, pp.71–78, Association for Computational Linguistics, 2003.
- [12] C.Y. Lin and F.J. Och, “Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics,” *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04)*, Main Volume, Barcelona, Spain, pp.605–612, July 2004.
- [13] E. DENOUAL and Y. LEPAGE, “Evaluation of translation using bleu in characters,” *Proc. of The 11th Annual Meeting of The Association for Natural Language Processing*, pp.522–525, March 2005.