

## ベクトル空間法を用いたイメージを想起させる お天気表現システム

飯田朱美、上野嘉人、松浦良平、相川清明

ake@media.teu.ac.jp, aik@media.teu.ac.jp

〒192-0982 八王子市片倉町 1404-1  
東京工科大学メディア学部

### 概要

本稿では気象情報などの多変量な環境情報からベクトル空間法を用いてイメージを想起させる表現を選択する環境表現システムを提案する。日常我々が天気などの環境情報を他者に伝える時に使っているような、情報の送り手受け手双方にとって感覚的に理解しやすい表現を収集し、環境情報表現リストを作成した。リストの各エントリーについて温度や湿度などの数値表現が可能な環境情報項目を設定し、各エントリーをそれらの項目に対する適合値のベクトル（テンプレートベクトル）として内部的に表現した。また、それらの合成ベクトルもあらかじめ計算し、合成ベクトルを含めてエントリーとした。そして外部から取り込んだ環境情報も同様にベクトル化し、入力ベクトルとテンプレートベクトルの類似度を計算し、類似度の最も高いエントリーを選択し出力した。また、類似度は一種の確信度と考えられるので、その値に応じて語尾表現を変化させた。実験システムを実装し評価を行った結果、提案手法を用いることで入力情報に合致した語句を選択できることがわかった。

## VECTOR-BASED WEATHER REPORTING SYSTEMS USING EMOTIVE EXPRESSIONS EVOKING SENDERS' METEOROLOGICAL IMAGERIES

Akemi Iida, Yoshito Ueno, Ryohei Matsuura, Kiyooki Aikawa

School of Media Science  
Tokyo University of Technology

### Abstract

This paper outlines our approach for making a computer select a qualitative expression to describe an environmental condition from given multivariate information using a vector-based method. First, we collected phrases we use daily to express environmental information and made a list of candidate phrases. By combining these phrases, we added more candidates to the list. The environmental conditions of the time which is given to the computer and the phrases in the list are both represented as vectors of quantitative data for environmental conditions such as temperature and humidity. By computing the similarity between the vectors of the given condition and those of each candidate phrase, the phrase that best represents the given condition is chosen. As an addition, we added a sentence-ending to the selected phrase which changes in accordance with the similarity value. Two systems have been developed to evaluate the validity of our method and results have shown that both systems successfully select suitable phrases for representing given environmental conditions.

## 1. はじめに

現在、気温や湿度などの環境情報の計測にコンピュータが用いられるのはごく一般的である。計測された値は様々なアプリケーションに渡され、画面にテキストとして出力されたり、音声合成を用いて出力されたりする。しかしながら、そのような表示は計測値の羅列であり味気ない。一方、人は環境情報をもっと感覚的に受け止め、様々な言い回しを使って表現する。例えば、季節は秋、時は10月のとある日、気温は摂氏24度、湿度30%、風力は穏やかな風でビューフォート風力階級3 [1]だとする。人は計算機ではないので実測はできないが、体感した気候をどう感じるかを簡潔な言葉で表現することができる。前述の例だとおそらく「さわやかだ」「すがすがしい」と表現する人が多いだろう。それらの表現を受け取った側も、おおむね、気温、湿度、風力などの送り側の環境を想起することができる。また、これらの表現は理解しやすいだけでなく、様々な環境条件を的確に伝えられるという観点から見ると効率的なコミュニケーションを可能にするものと捉えることもできる。一方、人は環境情報を比喩的な表現を用いて修辭的に表現することもある。例えば、吹雪いている状態を、川端康成は『二十年』 [2]で「降り落ちた雪もまた吹き上げられて、白い布のように飛ぶ」と表現している。我々の日常において、ここまで文学的に美しく表現する人は少ないだろう。しかし、我々も時には直接的な表現は避けて、環境情報に対する自己の印象を暗黙的に含めながら表現する場合もあるのではないだろうか。簡潔な表現も修辭的な表現も文面以上の行間を伝えようとしているという点で共通したものがある。人間のこのような言語行動をコンピュータにもさせようというのが本研究のコンセプトである。図1はこのコンセプトをイラストに表したものである。

本稿では我々のコンセプトを実現するために、ベクトル空間法を用いて、計測した気象情報などの多変量な情報を話し手聞き手双方にとって理解しやすい表現に置き換えて出力する手法を提案する。ベクトル空間法は情報検索の分野では、検索対象と質問文をキーワードのベクトルとみなし、その質問文と検索結果の類似度を求める手法として広く利用されている ([3] [4] [5])。実用的なシステムへの応用は、Zitouni, Kuo and Lee らの電話ルーティングシステム [6]などで検討されているが、現在のところ、我々の提案するようなコンセプト

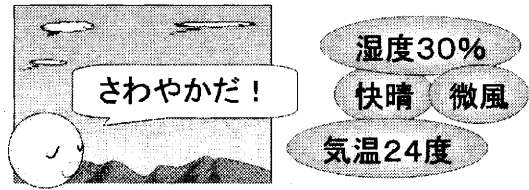


図1. システムコンセプト

にベクトル空間法を用いる研究はまだ報告がない。本稿では提案手法、実験システム、および、その評価について報告する。

## 2. 提案手法

本研究ではシステムが外部から取り込んだ環境情報と内部的に保持される環境情報表現リストをともに環境情報項目のベクトルとして表現した上で、二者の類似度を計算し最も類似度の高い語句を選択するという手法を提案する。ここでは前者を入力ベクトル、後者をテンプレートベクトルと呼ぶことにする。また、将来的には外部からの環境情報は各種センサーを用いて行う予定だが、今回の実験システムではグラフィカルユーザインタフェース (GUI) から入力した。

手順としては、まず日常用いられている環境情報を表現している語句を収集し、基本表現リストを作成した。収集した表現は定量的表現を使わずにその状態を情報の送り手聞き手双方に共通したイメージを想起させることができるものであり、感覚的な表現のものもあれば、比喩的なものもある。

次に気温や湿度などの数値表現が可能な環境情報項目を設定した。そして、基本表現リストの各エンタリ (基本エンタリー) をそれらの項目に対する適合値のベクトル (テンプレートベクトル) として表現した。テンプレートベクトルの適合値は、該当すれば1、該当しなければ0を基本としたが、ある特定のベクトル要素に重みを置く場合、1より大きい値を用いた。また、環境情報項目において、基本エンタリーの表現が複数の閾値に該当する場合には複数のベクトル要素において、最も該当する閾値を中心に徐々に変化する値を用いた。たとえば気温の場合には、 $-4^{\circ}\text{C}$  から  $38^{\circ}\text{C}$  まで2度刻みでベクトル要素を決定した。

また、一つの基本エンタリーを用いるよりも、二つ以上のエンタリーを組み合わせることにより、与えられた環境をうまく表現できることがある。例えば「べたつく」と「灼

表 1. 環境情報項目とベクトル要素

環境情報項目	風力			気温			湿度		
	1	3	6	10	20	30	30	60	90
さわやか	0	1	0	0	1	0	1	0	0
蒸し暑い	1	0	0	0	0	1	0	0	1
A	1	1	0	0	1	0	1	1	0

単位は次のとおり。風力: ビューフォート風力階級  
(0: 無風、1: 微風、3: 弱風、6: 強風)、気温: °C、湿度: %

熱」を連結させることにより、「ものすごく気温も湿度も高い」ということを表現できるわけである。そこで、複数の基本エントリーのベクトルを合成し、選択候補とした。原理的には基本エントリーの数までのベクトルの合成が可能であるが、今回は二つのベクトルの合成ベクトルまでとした。

その場合のテンプレートベクトルは二つの基本エントリーのテンプレートベクトルの論理和として求められるので、あらかじめ計算しておき、テンプレートベクトルとして登録しておく。従って、選択候補の語句の総エントリー数  $Sum$  は、基本エントリー数を  $N$  とした場合、式(1)で求められる。

$$Sum = N + \frac{N * (N - 1)}{2} \quad (1)$$

次に入力ベクトルと各エントリーのテンプレートベクトルの類似度は二つのベクトルの余弦として求めた。従って、 $i$  をベクトルの要素番号、入力ベクトルを  $x$ 、 $k$  番目のエントリーのテンプレートベクトルを  $v_k$  とした場合、 $x$  と  $v_k$  の類似度  $S_k$  (余弦) は式(2)で求められる。

$$S_{k,N} = \frac{\sum_{i=1}^N v_{ki} x_i}{\sqrt{\sum_{i=1}^N x_i x_i} \sqrt{\sum_{i=1}^N v_{ki} v_{ki}}} \quad (2)$$

ベクトルの要素は非負のため類似度の値は「0」から「1」の間であり、「1」に近い程、二つのベクトルの類似度が高いことを意味する。例として、図 2 に「さわやか」と「蒸し暑い」のベクトル要素の一部を示す。表 1 の中の「A」は入力情報のベクトルの一例である。

式(1)を用いて表 1 の入力ベクトル「A」と「さわやか」の類似度を求めると  $3/\sqrt{15}$  となる。同様に「A」と「蒸し暑い」の類似度を求める

と  $1/\sqrt{15}$  となり、後者の方が「1」に近い。従って、入力された環境条件には「さわやか」という表現が選択される。

図 2 は三次元ベクトル空間においてどのように表 1 の 2 つのエントリーから出力エントリーが選択されるかを図式している。

更に本研究では選択されたエントリーに語尾表現を追加して出力するようにし、この語尾の表現は、類似度の値によって表現を変えるようにした。類似度の閾値と語尾の語句は任意に設定が可能である。

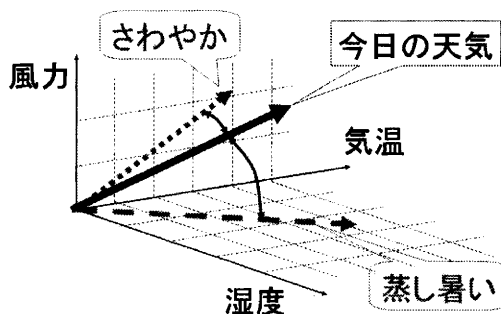


図 2. 三次元空間における提案手法の図式

### 3. 実験システム

提案手法の有効性を評価するために MATLAB 6.5.1 を用いて二つの実験システムを実装した。評価の結果、両システムは入力情報に合った表現を選択できることがわかった。

#### 3.1. Weather Reporter

Weather Reporter は 7 テーマの気象情報を取り込む。本システムは入力された数値情報をそのまま羅列するのではなく、その数値情報から相手に伝わりやすい簡潔な表現を選択し出力することを目的としている。

### 3.1.1. 環境情報語句リスト

文献[7][8]、および、Web ページから 149 の語や句を収集し、月、日、時間、天気、気温、湿度、風力の 7 つのテーマについて検討した結果、合計で 62 の環境情報項目に対する適合値のベクトルとして表現することにした。表 2 にその詳細を示す。

### 3.1.2. 合成ベクトル

Weather Reporter では二つのベクトルの合成ベクトルを生成した。合成ベクトルが最も類似度が高いベクトルとして選択された場合には、基本表現が二つ選択されたことになる。選択された二つの基本表現の間は接続詞「そして」で接続して出力した。表 3 は「暑い」と「ムシムシする」、そしてそれらを合成した「暑い、そして、むしむしする。」のベクトル要素である。

### 3.1.3. 確信度に応じて変化する語尾表現

Weather Reporter では、ヒューリスティックに類似度の閾値を設定し、1 に近づくほど確信度の高い語尾を用いた。選択された表現の品詞によって、追加する語尾を変形させる必要があったため、名詞、形容動詞、副詞を 1 群、動詞、形容詞を 2 群として処理を分けた。表 4 に類似度の閾値と語尾を示す。

### 3.1.4. 風景画像の取り込み

本システムでは CCD カメラからリアルタイムに風景画像を取り込めるようにした。起動時に風景画像を取り込んだ後、その画像は内部で RGB 値に解析され、三値を比較する。R の値が G、B より大きい時は本システム GUI の「時間」の選択肢は夜に設定され、B の値が R、G よりも大きいときには、「時間」が「昼」に、「天気」が「快晴」に設定される。

### 3.1.5. 画面表示

図 2 に Weather Reporter の表示画面を示す。画面の左上にはプルダウンメニューがあり環境情報項目の設定が行える。画面中央部には取り込んだ風景画像と RGB 値が表示される。画面右上の「SENSE」ボタンをクリックすると、システムは表現の選択処理を開始し、「QUIT」ボタンをクリックすると強制的に終了する。選択処理が終了すると画面下部の「VECTOR」欄には入力ベクトルが表示され、「RESULT」欄には選択された表現と語尾が出力される。本シ

テムはテキスト表示と合わせて音声合成による出力も行う。

表 2. Weather Reporter の環境情報項目

テーマ	環境情報項目	項目数
月	1月から12月	12
日	初旬、中旬、下旬	3
時間	午前、午後、夜	3
天気	快晴、晴れ、曇り、小雨、雨、雪、雷	7
気温	-4℃から to 38℃までの 2℃刻み	22
湿度	0 から 100% までの 10%刻み	11
風力	無風、微風、弱風、強風	4
合計		62

表 3. Weather Reporter の合成ベクトル

環境情報項目	気温 (°C)			湿度 (%)		
	10	20	30	30	60	90
暑い	0	0	1	0	0	0
蒸し暑い	0	0	0	0	0	1
暑い、そして、蒸し暑い	0	0	1	0	0	1

表 4. 類似度の範囲と語尾表現

類似度( $S_k$ )の閾値	語尾表現	
	名詞、形容動詞、副詞	動詞、形容詞
$0 \leq S_k < 0.4$	かな	
$0.4 \leq S_k < 0.45$	みたい	
$0.45 \leq S_k < 0.5$	らしい	
$0.5 \leq S_k < 0.55$	だと思う	と思う
$0.55 \leq S_k \leq 1.0$	だぞ	ぞ

## 3.2. Weather Poet

Weather Reporter は気象情報を表現する語句を気温や湿度を要素とするベクトルで表し、入力ベクトルに対して最も類似度が高い語句を選択するシステムだが、この考え方を修辭的な表現の選択にも応用できる。本研究では、その実現例として Weather Poet を作成した。

Weather Poet は、環境情報や送り手の感性情報を比喩的な表現を用いて表現するシステムである。比喩とは修辭法の一つであり、その中でも特に隠喩などはその前後の言葉との繋がりを保ちながら、直接的な物言いをせず、暗示的に表現することで、より読み手に訴える手法である。詳細な計測値を列挙する代わりに比喩を用い、相手のイメージの想起を期待することという点では効率的に情報を伝達

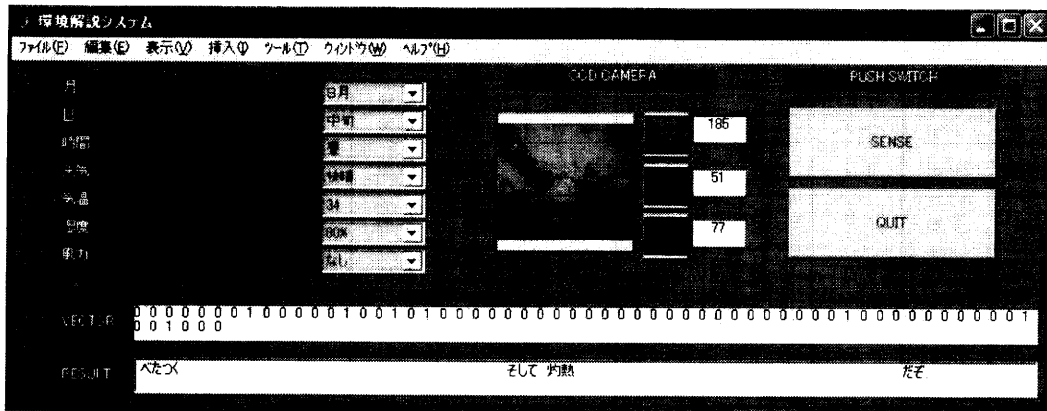


図 3. Weather Poet の表示画面

する方法の一つであると考えられ、その意味では Weather Reporter と共通するものがある。

本研究では[2]の中から 環境情報を含んだ比喩表現を 70 文登録した。第一節で取り上げた川端康成の文章もその中の一文だが、その外にも幾人かの小説家の文章を登録した。本研究ではそれらの比喩表現を 62 の環境情報項目に分類し、それらを環境ベクトルとした。入力画面ではこれらの環境情報項目を八つの大テーマからのプルダウンメニューから選択するようにした。八テーマとは季節、時間、天候、鉱物、環境、光・星、地形、感情である。今回の実装では一つ以上のテーマについて環境情報を設定すれば短文選択処理が実行されるようになっていた。例えば、入力画面で、「季節」のテーマから「冬」を、「時間」からは「夜」を、「光・星」からは「星」を選択すると、これもまた川端康成だが、『雪国』の「寒気が星を磨きだすように冴えて来た」という文章が出力される。

#### 4. 評価

両システムにそれぞれ 50 回、環境情報の入力を行い、評価を行った。表 5 は Weather Reporter が選択した語句の例である。全試行中、表 5 の三番目と五番目の合成ベクトル二例については、それを構成する二つの基本ベクトル同士に意味的な矛盾が見受けられるという点で、妥当な結果が得られなかった。Weather Poet では、より多くのテーマから環境情報を設定した方が、それらの情報を多く持つ短文を選択することができた。

#### 5. 今後の課題

まず Weather Reporter に関しては現行の実装では無条件にベクトルを合成するが、将来的には意味的に矛盾した基本エントリーの合成を避けるような工夫が必要であろう。また、今後は新たに検討する環境情報として、不快指数、気圧、および、過去の気象情報を考えている。前者二つは環境情報項目としての設定を検討する予定である。過去の気象情報については、天気予報でも「昨日よりは暖かでしょう。」や「一週間ほど前の肌寒いお天気に戻るでしょう。」という表現がよく使われることが挙げられることから、過去の気象情報のベクトルを情報として保持しておき、入力ベクトルとの比較に用いて、相対的な表現の生成を行いたいと考えている。天気予報に限らず我々も朝着る洋服を決める時に家族に相談すると「昨日は温かいから半袖のシャツを着れば良いんじゃない？」という返答が返ってくる場合も少なくない。このことから、相対的な表現は気象情報の自然な表現として重要だと考え、今後この機能についても実装していきたいと考える。

本稿で提案したコンセプトは基本的にはどの言語でも実装が可能である。しかし、今回の実装、特に Weather Poet に関しては、俳句や短歌など短い語句で環境情報や読み手の感性情報を表現する文化が根付いている日本だからこそ有効な結果を得られたということも可能性として考えられる。従って、他の言語で提案コンセプトを実現する場合はその言語が使われている国の文化を考慮して実装する必要があるだろう。

表 5. 入力ベクトルと選択された表現

	日	月	時間帯	天気	気温	湿度	風力	選択された表現	評価	類似度
1	下旬	1月	夜	雪	0°C	60%	強風	吹雪	適切	0.65
2	下旬	2月	朝	快晴	-4°C	30%	無風	身を刺す寒さ	適切	0.51
3	中旬	3月	夜	快晴	14°C	50%	強風	のどか、そして 身を切るような風	矛盾	0.55
4	下旬	6月	昼	雨	26°C	80%	無風	じと一つとした、 そして、梅雨空	適切	0.55
5	初旬	7月	朝	快晴	28°C	80%	無風	汗ばむ、そして、 雪起こし	矛盾	0.48
6	初旬	8月	昼	快晴	34°C	80%	弱風	灼熱、そして、ム ワツとした	適切	0.52
7	下旬	10月	昼	晴れ	30°C	80%	微風	べたつく、そし て、草いきれ	適切	0.54
8	初旬	11月	夜	快晴	24°C	10%	無風	からりと晴れる	適切	0.49

評価：適切：適切なエントリーが選択された。

矛盾：合成ベクトルを構成する基本エントリー同士が意味的に矛盾している。

## 6. 結論

本稿では気象情報などの多変量な環境情報からイメージを想起させる表現を選択する環境表現システムを提案した。提案手法の特徴はベクトル空間法を用いたことにある。まず環境情報を感覚的に理解しやすい語句で表した表現のリストを作成した。そして、リストの各エントリーについて気温や湿度などの数値表現が可能な環境情報項目を設定し、それらの項目に対する適合値のベクトルとして表現した。入力ベクトルと各エントリーのテンプレートベクトルの類似度を計算し、類似度の最も高いエントリーを選択し出力した。また、類似度の値に応じて語尾表現を変化させた。実験システムを実装し評価を行った結果、提案手法を用いることで入力情報に合致した表現を選択できることがわかった。本手法は、五感センサーを持つロボットや擬人化エージェントと人の効率的な対話に有効であると考えられる。今後はよりの確で自然な表現を出力するために、感情表現リストの洗練、意味的に矛盾のない合成ベクトルの生成、および、過去の環境情報との相対的な表現の生成について検討する予定である。また、語尾表現を変えるための類似度の閾値についてもさらに検討を行う。また、ベクトルを合成する時に、現在は全て「そして」という接続詞で連結させているが、より自然な連結についても検討を行う予定である。

## 7. 謝辞

確信度を反映した言語表現について討論していただいたバイオニクス学部三田地成幸教授、高野光雄助手に感謝致します。また日頃、ご助言いただきますメディア学部飯田仁教授、武村幸雄助手、大庭真人助手に感謝致します。

## 8. 参考

- [1] National Weather Service Chicago, the Beaufort Wind Scale (n.d.). [http:// www.crh.noaa.gov/lot/webpage/beaufort/](http://www.crh.noaa.gov/lot/webpage/beaufort/)
- [2] 中村明、『比喩表現辞典』、角川書店、1995.
- [3] Salton, G. "Developments in automatic text retrieval". Science 253, 974-980, 1991.
- [4] Jurafsky, D., Martin, J. "Speech and language processing: an introduction to natural language processing", in: Computational Linguistics, and Speech Recognition. Prentice Hall, 2000.
- [5] Kuo, H. -K. J. and Lee, C.-H. "A portability study on natural language call steering", In: Proceedings of the Eurospeech-01, Aalborg, Denmark, 2001.
- [6] Zitouni, I., Kuo, H. -K. J. and Lee, C.-H. "Boosting and combination of classifiers for natural language call routing systems", Speech Communication 41, 647-661, 2003.
- [7] 日本放送協会編、『最新気象用語ハンドブック』、日本放送協会(NHK ブックス)、1986.
- [8] 気象庁ホームページ (n.d.). from [http://www.jma.go.jp/JMA\\_HP/jma/index.html](http://www.jma.go.jp/JMA_HP/jma/index.html)