A VECTOR-BASED METHOD FOR EFFICIENTLY REPRESENTING MULTIVARIATE ENVIRONMENTAL INFORMATION

Akemi Iida, Yoshito Ueno, Ryohei Matsuura, Kiyoaki Aikawa

School of Media Science
Tokyo University of Technology
ake@media.teu.ac.jp, aik@media.teu.ac.jp

ABSTRACT
This paper outlines our approach for describing multivariate environmental information such as weather as it might be characterized by humans using connotations and delicate nuances. The purpose of this research is to achieve smooth human-machine spoken dialogue. The key feature of our approach is the use of a vector-based method, a widely used technique in Information Retrieval research. To date, no system concept as mentioned or the use of the vector-space in such a system has been reported. First, phrases implicitly expressing environmental information are collected to make the list of candidate phrases. Next the environmental conditions of that time are entered from the GUI. The phrases and the newly entered information are represented as vectors of component values representing such information. By computing the similarity (relatedness) between the given condition and each candidate phrase, the phrase that best represents the given condition is chosen. Two experimental systems that work in Japanese have been developed to evaluate the validity of our method. The first system selects the appropriate phrase to represent the weather implicitly with a minimum number of phrases and not by lining up measured values. The other system selects poetic phrases that most appropriately represent the environmental conditions, including weather information and emotional states. The result of our evaluation has shown that both systems successfully selected the suitable phrases for representing given environmental conditions.

1. INTRODUCTION
With recent developments, computers are now widely used for sensing temperature, humidity and other environmental conditions. The results are recorded and used for various applications and in some systems are announced by text and speech synthesis. However, such outputs are extremely colorless displays or readings of measured figures that are very insipid. For autonomous agent systems, we consider that a more desirable description should be more expressive, as people are sensitive to environmental conditions and can describe how they feel with various expressions. For example, let us take a day in October, when the temperature is 24° centigrade, humidity is 30%, and the wind power is 3 using the Beaufort Wind Scale (gentle breeze) [1]. Instead of listing those figures, people may refer to such weather as ‘bracing’, ‘refreshing’, or ‘invigorating’. This example shows that humans can condense measured values the system sensed into a single word. Another example is the reverse of this. Let’s say that the season is the end of autumn and a person feels wistful or sentimental about the passing of autumn. He might want to express how he feels by a literary quotation such as “The dragonfly is so weak that it is almost as though you can see right through him.” This paper describes our novel approach of using a vector-based model to make the computer represent environmental conditions as humans might just like the examples given above. In the following section, some of the related studies are introduced, followed by a description of our approach, our experimental systems, and evaluation of our approach.

2. RELATED STUDIES
To date, the use of the vector-space for realizing a system concept such as that outlined above has not been reported. The vector-based method is a widely used technique in Information Retrieval research. There, for every topic in the training corpus, queries are represented as vectors of component values representing the frequency of terms that occur within them and the distance between the query vector and each topic vector is computed. The classifier then selects the topic that is the closest to the query ([2] [3] [4]). Zitouni, Kuo and Lee’s research [5] on call routing systems is a good application-oriented example that uses such a method.

3. THE APPROACH TAKEN IN THIS RESEARCH
First, phrases that describe environmental conditions are collected. Each phrase is then represented as a vector of ‘the goodness of fit’ for vector components such as temperature and humidity. Hereafter, each vector is referred to as a ‘phrase vector’ and the set of phrase vectors is referred to as a ‘vector table’. As standard setting, two values are provided for the goodness of fit; ‘present’ represented as 1 and ‘absent’ represented as 0.

The goal of our approach is, using the sensors, to have the system sense the environmental conditions and to form a vector (hereafter, the ‘query vector’) following the same manner applied to form phrase vectors. The experimental systems utilize a GUI for the user to select information to set the goodness of fit for each vector component. The similarity between the query vector and each phrase vector is computed by the following procedure: First the norm for each vector is computed and then each component value is divided by the norm to obtain the normalized value. Then the cosine similarity measure derived from the dot product of the query vector and each phrase vector is computed and compared. The cosine obtained for the two vectors corresponds to the similarity between the two. The pair which has the closest
cosine value to 1 can be referred to as the pair with the highest degree of similarity.

Table 1 shows how each phrase vector holds values giving vectors for ‘bracing’ and ‘muggy’ as examples. The query vector is also listed in Table 1 represented as ‘A’ for illustrative purposes.

<table>
<thead>
<tr>
<th>Category</th>
<th>Wind</th>
<th>Temperature</th>
<th>Humidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bracing</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Muggy</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

The unit for temperature is centigrade (°C), for humidity is percentage (%), for wind velocity is Beaufort Wind Scale (1: light air, 3: gentle breeze, 6: strong breeze) [1].

With the Eq. (1) where \( i \) denotes the component index of the vector, the similarity \( S_k \) (the cosine similarity measure) between the query vector \( x \) and the \( k \)th phrase vector in the vector table \( V_k \) can be computed.

\[
S_k = \frac{\sum_{i=1}^{N} V_{ki} X_{i}}{\sqrt{\sum_{i=1}^{N} X_{i}^2} \sqrt{\sum_{i=1}^{N} V_{ki}^2}}
\]

Eq. (1)

Fig. 1 depicts schematically explanation in a three-dimensional vector space how the appropriate phrase is selected using the vectors listed in Table 1. Applying Eq. (1) to the given example, the cosine measure, \( S_b \) between the query vector ‘A’ and ‘bracing’ as shown in equation 2 is closer to 1 than that of ‘A’ and ‘muggy’, \( S_m \) as in equation 3.

\[
S_b = \frac{3}{\sqrt{15}}
\]

Eq. (2)

\[
S_m = \frac{1}{\sqrt{15}}
\]

Eq. (3)

Hence, today’s weather is represented as ‘bracing’ in our approach.

4. EXPERIMENTAL SYSTEMS

Two systems that work in Japanese have been developed using MATLAB 6.5.1 to evaluate the validity of our approach. The first system selects a suitable phrase to represent the multivariate weather efficiently and expressively. The other selects poetic phrases that most appropriately represent the environmental conditions and the users’ emotional states.

4.1. The Weather Reporter

The Weather Reporter measures various weather-related values. Instead of listing all those values, it aims to express the multivariate weather information efficiently and expressively with a minimum number of phrases and not by listing all of the various measured values.

4.1.1. List of Phrases

149 phrases are collected from [6], and Japan Meteorological Agency’s website [7] and other resources on the internet. Further, 51 idiomatic phrases are added and a total of 200 phrases are then represented as a vector of the goodness of fit for 62 vector components as shown in Table 2. Values are in

Figure 1: Schematic diagram of our selection method

Figure 2: The Weather Reporter’s GUI
principle 0 or 1 as stated in Section 3, but for some phrases that are typically used to represent particular conditions some vector values are weighted heuristically so that the phrase will be selected.

4.1.2. Combined Vectors

One interesting function of the Weather Reporter is the generation of a ‘combined vector’. This function enables the system to take the value for logical ‘OR’ of two vectors and generate the combined vector. Below is an example with ‘hot’ and ‘humid’ which generates the expression, “hot and humid”. By expanding this function to combine three or more vectors, the system can generate more complicated expressions and can convey more delicate nuances.

4.1.3. Capturing the environmental image

In addition to the input method described above, the weather reporter facilitates the use of a CCD camera that enables the system to capture the environmental image real-time. The system decomposes the image to RGB values. Each value for R, G, and B is then compared with each other. When the value for R is larger than those of G and B, the system sets the default value for the time of day as ‘night’. Likewise, when the value for B is larger than those of R and G, the system sets the default value for the time of day to ‘daytime’ and the weather to ‘clear’.

4.1.4. Graphical User Interface

Figure 2 shows the graphical user interface (GUI) for the weather reporter. The value selection window is located at the top left and to its right is a real time image of the environment captured by the CCD camera. Rightmost are the push switches, ‘sense’ and ‘quit’; the former is to activate the computation and the latter is to cancel and exit the system. Beneath the three components just explained a window displaying the vector values given from the value selection window is located. At the very bottom, there is a window displaying the selected phrase or phrases. In addition to the graphical output, the weather reporter reads out the phrase/phrases with the speech synthesizer.

4.2. The Weather Poet

The Weather Poet is the other system that has been developed. It has the same capability as the weather reporter. Two features in particular should be mentioned for this system. The first feature is that the Weather Poet selects values not only for weather-related conditions but also for other environmental conditions such as landforms, lights and stars, and mineral substances such as sands, stones, metals, and most notably, the users’ emotional states. Thus the system has altogether 62 vector components. The second feature of note is that the Weather Poet outputs poetic phrases that most appropriately represent the given conditions. The phrase is selected from 70 metaphoric phrases chosen from [8]. An example of a selected poetic phrase is given in the Introduction; the Weather Poet is given the information that represents the end of autumn and the information that the user feels wistful or sentimental about the passing of autumn. Using the proposed method, the system outputs a poetic phrase such as “The dragonfly is so weak that it is almost as though you can see right through him.”

5. EVALUATION

For both the Weather Reporter and the Weather Poet, evaluation was conducted by the authors with 50 different conditions given to each system. Table 4 shows selected phrase or phrases for the Weather Reporter. Among all the trials, only two examples shown as index 3 and 5 show some contradiction between the two phrases generated by combining vectors (cf. 4.1.3). For the Weather Poet, precision improves when more vector components are filled in with values.

6. DISCUSSION

Regarding the Weather Reporter, two things are suggested for improvement. We need to consider how to avoid selecting a combined vector of contradictory pairs. Currently the system automatically generates combined vectors without any restriction but for our future work, some restriction module may be necessary. Another thing to be considered is to increase the number of vector components to improve precision in selection; the candidate components are the temperature-humidity index, barometric pressure and the previous day’s weather.
The next topic is related to both systems. It is important to pay attention to different methods of expression when the culture is different. The implementation of our approach worked well in Japanese since we are people who have a long history of expressing weather and other environmental conditions with a variety of different phrases and we are also keen in representing them with our emotion reflected in short verses such as *haiku* (17-syllabled verse which always includes a seasonal phrase). However, this might not apply to other cultures and therefore cultural differences must be taken into account when developing such a system as described in this paper.

### 7. CONCLUSION

An approach for efficiently representing multivariate environmental information by short phrases has been described. The key feature of our approach is the use of a vector-based method. The phrases in the phrase table and newly-entered environmental conditions are represented as vectors of component values implicitly representing such information, and the similarity between the condition given and each candidate phrase is computed to select the most appropriate phrase describing the condition. Two systems that work in Japanese have been developed and the result of our evaluation has shown that both systems selected suitable phrases successfully and that our approach is valid.

### 8. ACKNOWLEDGEMENTS

The authors would like to thank Prof. Hitoshi Iida, Prof. Seiko Mitachi, Dr. Yukio Takemura and Mr. Masato Ohba for valuable discussions and also to Prof. Paul Brocklebank for proofreading.

### 9. REFERENCES


---

**Table 4: Given condition and selected phrases**

<table>
<thead>
<tr>
<th>Part of the month</th>
<th>Month</th>
<th>Part of the day</th>
<th>Weather</th>
<th>Temp.</th>
<th>Hum.</th>
<th>Wind</th>
<th>Phrase(s)</th>
<th>Eval.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Late</td>
<td>Jan.</td>
<td>Night</td>
<td>Snow</td>
<td>0°C</td>
<td>60%</td>
<td>Strong breeze</td>
<td>Fubuki</td>
</tr>
<tr>
<td>2</td>
<td>Late</td>
<td>Feb.</td>
<td>Morning</td>
<td>Clear</td>
<td>-4°C</td>
<td>30%</td>
<td>No Wind</td>
<td>Miwosasu samusa</td>
</tr>
<tr>
<td>3</td>
<td>Mid</td>
<td>March</td>
<td>Night</td>
<td>Clear</td>
<td>14°C</td>
<td>50%</td>
<td>Strong breeze</td>
<td>Nodoka &amp; Miwkiruyouna kaze</td>
</tr>
<tr>
<td>4</td>
<td>Late</td>
<td>Jun.</td>
<td>Daytime</td>
<td>Rain</td>
<td>26°C</td>
<td>80%</td>
<td>No Wind</td>
<td>Jitootoshita &amp; tsuyuzora</td>
</tr>
<tr>
<td>5</td>
<td>Early</td>
<td>July</td>
<td>Morning</td>
<td>Clear</td>
<td>28°C</td>
<td>80%</td>
<td>No Wind</td>
<td>Asaburu &amp; Yukikosho</td>
</tr>
<tr>
<td>6</td>
<td>Early</td>
<td>Aug.</td>
<td>Daytime</td>
<td>Clear</td>
<td>34°C</td>
<td>80%</td>
<td>Gentle breeze</td>
<td>Shakanetsu &amp; Muwattoshita</td>
</tr>
<tr>
<td>7</td>
<td>Late</td>
<td>Oct.</td>
<td>Daytime</td>
<td>Clear</td>
<td>30°C</td>
<td>80%</td>
<td>Light air</td>
<td>Betatsuku &amp; Kusaikire</td>
</tr>
<tr>
<td>8</td>
<td>Early</td>
<td>Nov.</td>
<td>Night</td>
<td>Clear</td>
<td>24°C</td>
<td>10%</td>
<td>No Wind</td>
<td>Kararito hareru</td>
</tr>
</tbody>
</table>

English Translation: 1) blizzard, 2) bitterly cold, 3) moderate but penetrating wind, 4) muggy with clouds typical of the rainy season, 5) sticky and thundery just before a snowfall, 6) severely hot and humid, 7) oppressively sticky with high grass pollen count, 8) crisp, clear skies.

Abbreviations: Temp.: Temperature, Hum.: Humidity, Wind: Wind Velocity, Eval.: Evaluation, OK: Selection was appropriate, NG (Not good): Selection was inappropriate.