Towards Voice-Input Symbolic Pattern Retrieval using Parameter-Based Search

Yukiko Suzuki ¹, Kiyoaki Aikawa ¹

¹School of Media Science, Tokyo University of Technology, Tokyo, Japan
bovo444@gmail.com, aik@media.teu.ac.jp

Abstract
This paper proposes a symbolic pattern retrieval method using emotional feature vectors. Queries and symbolic patterns are represented by emotional vectors composed of eight numerical parameters. Since the proposed method uses numerical vectors close to raw data instead of recognized text, the information loss by data conversion is small. This point is advantageous compared with conventional text-based search such as recent spoken document retrieval approach. Five similarity measures were compared on a test collection. The cosine similarity and the Euclidean distance showed the best performance among five similarity measures. OOV analysis clarified several problems for achieving voice-input symbolic pattern retrieval.

Index Terms: pattern retrieval, emotional feature, parameter-based, similarity measure, voice-input

1. Introduction
Spoken document retrieval is getting important in information retrieval. One of the authors has been collaborating with several research organizations on building up test collections for spoken document retrieval [1-5]. A special session on the topic was organized by these researchers at INTERSPEECH2010 [6]. One reason of the importance is explosively increasing raw audio-visual data located inside PCs and on the WWW. Movie and speech files are included in the audio-visual data. The other reason is the progress of high performance automatic speech recognition (ASR). Once the speech data are transcribed into text using ASR, the conventional text search techniques are available. This approach is called the spoken document retrieval. However, automatic recognition is not available for most kinds of raw data. Symbolic pattern is one of such data. A problem exists even for the recognizable speech data. If the utterance is unclear or the speech is contaminated by noise, some keywords may be lost or misrecognized by ASR. The incorrect transcription results in degrading search performance.

On the other hand, emotional feature-based music retrieval was proposed by one of the authors [7-10]. The query and music pieces were given by eight-dimensional emotional feature vectors. Each feature is represented in 5-level value (brightness=5, sadness=1, etc). The music retrieval system search music using the similarity between the query and music vectors. This work was preceded by the weather information tags or metadata are not required for search. However it is difficult to apply such recognition-based method into symbolic pattern search, because symbolic pattern cannot be uniquely transcribed into text. This paper proposes a parameter-based search for overcoming the disadvantages of text-based search using automatic recognition. Figure 1 compares conventional text-based search and proposing parameter-based search. The numerical parameter vector can preserve the information of raw data, while some information is lost by data-to-text conversion. The proposing approach is different from conventional matching-based data scanning or spotting in the sense that parameters are derived for both of data and queries.

The proposing parameter-based method is characterized by the search on the feature parameter instead of text. This method is advantageous on the following points.
(a) The proposing method can search objects which are lost by automatic recognition.
(b) The method can repeatedly search weak features by decreasing the detection threshold. The detection threshold may be controlled by queries such as "Search more".

Figure 1: Comparison of text-based search and parameter-based search.

The section 2 describes the parameter-based search method. Section 3 describes the symbolic pattern search using emotional features. Section 4 derives similarity measures between parameter vectors. Section 5 shows the experimental method and results. Section 6 concludes this paper.
The method can modify query vectors according to the additional query. The data bias such as microphone transfer function can be compensated by the additional query. Figure 2 illustrates the mechanism of query modification for vector-based search. It also enables the synthesis of statistical models like PMC for noisy speech recognition [13].

**Table 1. Example of the emotional levels for query terms.**

<table>
<thead>
<tr>
<th>Emotional Feature</th>
<th>Query Term</th>
<th>Complex</th>
<th>Soft</th>
<th>Strong</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Square</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Smooth</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Rough</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Simple</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Example of the emotional levels for symbolic patterns.**

<table>
<thead>
<tr>
<th>Emotional Feature</th>
<th>Pattern</th>
<th>Complex</th>
<th>Soft</th>
<th>Strong</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3. Examples of probability density functions of emotional features for symbolic patterns. The horizontal axis is the emotional level.**

**3. Pattern Search using Emotional Feature**

This paper is an attempt of retrieving symbolic patterns shown in Fig. 3. The goal of this study is voice-input retrieval. This paper used emotional representation which was natural for users to specify symbolic patterns in queries. Eight emotional features were selected for searching the symbolic patterns. The selected features were "complex", "bright", "soft", "warm", "light", "vague", "strong", and "energetic". A test collection for evaluating search method was manually created. Queries for evaluating the retrieval system were collected for searching the 100 symbolic patterns shown in Fig. 3. The emotional parameter data were obtained for the 58 queries collected by subjective experiments. The subjects were requested to answer 5-level value for the eight emotional parameters. Number of subjects was 20. Then, the probability distribution of level was obtained for each of eight parameters query by query. The emotional vectors for symbolic patterns were also obtained by the same procedure.

**4. Similarity Measure**

Both of query terms and symbolic patterns are represented by emotional vectors. Table 1 and Table 2 show the examples of emotional levels for queries and symbolic patterns, respectively. Here, three of eight emotional features are shown. A query was composed of one or more query terms. As for the symbolic patterns, the emotional levels were collected over numbers of subjects. Therefore, both of the query and symbolic pattern vectors had probability distribution in emotional level axis. Table 3 schematically illustrates the example of probability distribution of emotional levels. Symbolic pattern search can be performed using the similarity between a symbolic pattern vector and the query vector. The query vector was obtained by averaging vectors corresponding the query terms. This paper compared five similarity measures on search performance. Likelihood, PDF (probability density function) distance and Kullback-Leibler divergence used probability distribution of emotional levels. Cos similarity and Euclidean distance used the mean of each probability distribution. The distances and the divergence indicated dissimilarity.
4.1. Likelihood

The PDFs of emotional levels are approximated by Gaussian Mixture Models (GMMs). A GMM is composed of five Gaussian distributions each centered at 1 to 5 in the level axis. The likelihood of GMMs to the given query vector is used for the similarity measure.

\[ p_n(i) = \sum_{j=1}^{5} p_{n}(i) \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(i-j)^2}{2\sigma_n^2}\right), \]

where \( N \) is the vector size and \( L \) is the number of levels of emotional parameters. In this paper \( N=8 \) and \( L=5 \). The variance of each Gaussian is normalized to 1.

4.2. PDF Distance

This measure represents the distance between two PDFs of a query and a symbolic pattern. The measure is defined by the expectation of Euclidean distance of emotional level difference. The mean and variance of the probability distribution of the \( k \)-th emotional parameter of the \( n \)-th symbolic pattern and those of \( k \)-th emotional parameter of the query are given by

\[ \mu_{nk} = \sum_{i=1}^{L} p_{nk}(i) \mu_k(i), \]

\[ \nu_k = \sum_{i=1}^{L} i q_{nk}(i), \]

\[ \sigma_{nk}^2 = \sum_{i=1}^{L} p_{nk}(i) \mu_k(i) - \mu_{nk}^2, \]

\[ \rho_k^2 = \sum_{i=1}^{L} i^2 q_{nk}(i) - \nu_k^2. \]

The PDF distance is given by the following equation.

\[ d_{nk} = \sum_{i=1}^{L} \sum_{j=1}^{L} p_{nk}(i) q_{nk}(j)(i-j)^2 = \sum_{i=1}^{L} p_{nk}(i) q_{nk}(j)(i^2 - 2ij + j^2), \]

\[ = \sigma_{nk}^2 + \mu_{nk}^2 - 2\nu_k \mu_{nk} + \left(\mu_k^2 + \nu_k^2\right) \]

\[ = \sigma_{nk}^2 + \rho_k^2 + \left(\mu_{nk} - \nu_k\right)^2. \]

The distance between the query and the \( n \)-th symbolic pattern is obtained by the sum of the Eq. (6) over whole emotional parameters as

\[ d_n = \sum_{k=1}^{N} d_{nk}. \]

4.3. Kullback-Leibler Divergence

From the view point of information amount, the Kullback-Leibler divergence is appropriate for measuring the difference of two PDFs. The symmetric Kullback-Leibler divergence between the PDFs \( p_n(i) \) and \( q_n(i) \) is given by

\[ V_n = \sum_{k=1}^{N} \left[ \sum_{i=1}^{L} p_{nk}(i) \log \frac{p_{nk}(i)}{q_{nk}(i)} + \sum_{i=1}^{L} q_{nk}(i) \log \frac{q_{nk}(i)}{p_{nk}(i)} \right]. \]

4.4. cos Similarity

The cos similarity uses only the mean of emotional level distributions. The cos similarity is calculated using the dot products of two mean vectors. The mean of the \( k \)-th emotional parameter of the \( n \)-th symbolic pattern is given by Eq. (2). The mean of the \( k \)-th emotional parameter of the query is given by Eq. (3). Then, the cos similarity is given by

\[ c_n = \frac{\text{dot} (\mu_{nk}, \nu_k)}{\sqrt{\text{dot} (\mu_{nk}, \mu_{nk}) \text{dot} (\nu_k, \nu_k)}} = \sum_{k=1}^{N} \mu_{nk} \nu_k \sqrt{\sum_{k=1}^{N} \mu_{nk} \mu_{nk} \sum_{k=1}^{N} \nu_k \nu_k}. \]

4.5. Euclidean Distance

This measure also uses only the mean of emotional level distributions. The Euclidean distance between the emotional vector of the query and that of \( n \)-th symbolic pattern is given by

\[ D_n = \sum_{k=1}^{N} (\mu_{nk} - \nu_k)^2. \]

5. Experiments

5.1. Method

In symbolic pattern retrieval, more than one correct candidate may exist for the given query. Then, three best candidates were manually selected as the primary correct candidates for each query. Up to five secondary acceptable candidates were also manually selected. The number of secondary correct candidates was different among symbolic patterns. The evaluations on five similarity measures were carried out using these primary and secondary correct candidates.
5.2. Comparison of Five Similarity Measures

Table 4 shows the number of retrieved symbolic patterns for three detection thresholds. Figure 4 compares similarity measures on recall and precision. Figure 4 (a) shows the results for primary correct candidates and Figure 4 (b) shows the results for primary and secondary correct candidates. The results indicated that similarity measures show almost the same performance excepting likelihood. The best precision for the top candidate was 0.76, which was obtained by the cosine similarity and the Euclidean distance. The results suggested the controllability of detection sensitivity by adjusting thresholds by queries.

![Figure 4: Comparison of five similarity measures on symbolic pattern retrieval performance.](image)

6. Conclusions

This paper proposed a parameter-based search method for symbolic pattern retrieval. The purpose of this method was the retrieval of raw data which was difficult to be transcribed into texts using automatic recognition. This paper used emotional feature parameters which was natural for users to ask queries. Five similarity measures were compared on search performance. The cosine similarity and the Euclidean distance showed the best performance. OOV analysis suggested that the improvement of mimetic word recognition was necessary for achieving voice-input symbolic pattern retrieval.

7. Acknowledgements

This paper used Japanese ASR engine Julius 4.1.3 and the dictation kit.

8. References


5.3. OOV test

Additional 92 queries were collected for analyzing OOV terms for voice-input retrieval. Then, the total number of tested queries was 150. The OOV rates were examined using the automatic recognition. This paper used emotional feature parameters which was natural for users to ask queries. Five similarity measures were compared on search performance. The cosine similarity and the Euclidean distance showed the best performance. OOV analysis suggested that the improvement of mimetic word recognition was necessary for achieving voice-input symbolic pattern retrieval.

Table 5. OOV terms in queries.

<table>
<thead>
<tr>
<th></th>
<th>58 query set</th>
<th>92 query set</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>OOV terms</td>
<td>22</td>
<td>56</td>
<td>78</td>
</tr>
<tr>
<td>OOV rate</td>
<td>0.38</td>
<td>0.61</td>
<td>0.52</td>
</tr>
</tbody>
</table>