Lecture subtopic retrieval by retrieval keyword expansion
using subordinate concept

Noboru Kanedera 1, Tetsuo Funada 2, Seiichi Nakagawa 3

1 Ishikawa National College of Technology, Japan
2 Kanazawa University, Japan
3 Department of Information and Computer Sciences, Toyohashi University of Technology, Japan

kane@ishikawa-nct.ac.jp, t_funada@aqua.plala.or.jp, nakagawa@slp.ics.tut.ac.jp

Abstract
We developed a supporting system for creation of educational video contents. The system automatically segments a lecture video material into subtopics based on speech signals by a statistical model for text segmentation. In this paper, we report on the result of retrieving the lecture subtopics by keyword expansion using the knowledge of the dictionary, and so on. The keyword expansion using the subordinate concept improved the average reciprocal order (MRR: Mean Reciprocal Rank) from 0.51 to 0.55 when subtopics are retrieved by a set of three search keywords for the lecture voice text recognized by automatic speech recognition.

Index Terms: subtopic retrieval, keyword expansion, automatic speech recognition, lecture video

1. Introduction
The method of retrieving a subtopic of a lecture video is examined. When the subtopic of the lecture video is retrieved, we can use slide information [1], etc. However, only speech information is used in this study, because the lecture using blackboards that don't use the slide is targeted.

When speech recognition is used, measures against the speech recognition errors are important. The method of using the subword [2][3][4], a statistical translation technique [5], etc. have been proposed. The methods based on query expansion using related web documents [6][7] and corpus [8] have been also proposed.

In this paper, we report on the result of investigating which relationship is important among hypernym and hyponym relationships in retrieval keyword expansion.

2. Lecture subtopic retrieval

2.1. Retrieval keyword expansion
In the text retrieval, the keyword expansion might be used. This study investigates how the retrieval keyword expansion is effective for the transcription text of lecture speech and the text by automatic speech recognition.

When the method of retrieval keyword expansion is classified in the source, the method of using the dictionary, the method of using information such as Web, and the method of using together, etc. are devised. A general dictionary doesn't often contain the technical term though it contains a lot of general terms. Oppositely, the possibility including the technical term is high, though the method of using information such as Web might not contain a general term.

The broader concept, the subordinate concept, and the synonym, etc. are considered when classifying it by the concept. We investigated the method to use Japanese Wordnet [9] which can retrieve these concepts.

Figure 1 shows the method of retrieving the subtopic of lecture video. The index (TF-IDF) of each subtopic of the lecture video is obtained beforehand. When retrieving it, the specified keywords by a user are expanded to the association key words such as broader concepts, subordinate concepts, and simultaneous appearance words, etc. using the dictionary [10] and Wordnet. Next, the retrieval vector is generated. The constant value $\alpha$ is set to the retrieval vector element corresponding to the retrieval keywords. The value $(1 - \alpha) / (\text{the number of associated keywords})$ is set to the retrieval vector element corresponding to the association keywords. In this report, $\alpha$ was assumed to be 0.25. The retrieval is done by comparing the retrieval vector with the TF-IDF vector of each subtopic.
2.2. Subword Retrieval

The keyword expansion can retrieve the subtopic to which the keyword doesn't appear directly by expanding the specified keyword. On the other hand, the method using the subword can deal with OOV (out of vocabulary)[4]. It is therefore preferable to use both the keyword expansion and the method using the subword.

Figure 2 shows a lecture subtopic retrieval method by subword model. Tri-phone was used as a subword. Confusion matrix between tri-phone models is made. Bhattacharyya distance [3] was used for the distance between distributions of tri-phone models. The uttered keywords are converted into the tri-phone string, and collated with the tri-phone strings of subtopics by continuous dynamic programming (DP) using the confusion matrix. As a result of continuous DP, N sections where the spotting distance is small are extracted. In this report, N was assumed to be 500. N sections are sorted in order with small distance. The score of each section is defined as 1/sorted rank, and it adds to the corresponding subtopic. The subtopic with a large score was assumed to be a subtopic retrieval result.

3. Experiment

3.1. Experimental data

Table 1 shows the used lecture videos uttered by five male teachers. Each class video is approximately 90 minutes. The subject names of the videos are operating system, mathematics of information engineering, digital signal processing, digital circuit, and communication engineering. Video materials 1, 4, and 5 are lectures that used the presentation software. Video materials 2 and 3 are lectures with a blackboard. The number of sentences in Table 1 is the number of speech intervals divided by whether or not a silent interval continued for 1 second or more. As the recording was made using a headset, there is little effect due to noise, etc. Speech information alone was extracted from these videos, and down-sampled at 16 kHz.

Table 2 shows speech recognition results for the lecture videos. The recognizer was Julius-4.0[11]. The Linguistic model was trained from the Corpus of Spontaneous Japanese (CSJ) [12][13]. The acoustic model was a tri-phone model (3000 states, 16 mixtures) trained from lecture speech in academic meetings. Automatic speech recognition is very difficult as shown in Table 2, because the perplexity (defined as the entropy's power of 2 ; possible candidate number)[14] of the lecture is high.

The correct subtopic boundaries were determined as follows: i) The tolerance ranges of subtopic boundaries for the lecture video were extracted by five evaluators. ii) The tolerance ranges, where three evaluators or more are consistent, were extracted. iii) The center of the tolerance range was defined as a correct boundary. These subtopic boundaries were used in the following experiments.

3.2. Evaluation results

Figure 3 shows the retrieval results of subtopics of lecture videos. The vertical axis is MRR (Mean reciprocal rank)[11]. In Figure 3, ‘Ideal’ means the retrieval results using ideal keywords selected from the transcription text of each subtopic in order in which TF-IDF is large. ‘Real’ means the retrieval results using written keywords extracted by five evaluators.

The retrieval performance has worsened in the real keyword in compared with ideal keyword. For example, MRR has decreased to 0.677 in the actual keyword while MRR is 1.0 in ideal keyword when the transcription text is used, and a set of three keywords for each query is used. When the text by ASR and three keywords are used, MRR has decreased to 0.511 in the actual keyword while MRR is 0.802 in ideal keyword. This result shows that measures against the out-of-vocabulary, etc. are necessary.

Figure 4 shows how keyword expansion is effective for ASR text. The retrieval keywords are extracted by five evaluators. In Figure 4, ‘Dic’ means the results with keyword expansion by a dictionary. ‘Wordnet (subordinate)’ means the results using the subordinate concept of the retrieval keyword

<table>
<thead>
<tr>
<th>Lecture video</th>
<th>The number of sentences</th>
<th>The number of correct boundaries</th>
<th>The number of subtopics</th>
<th>Perplexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>539</td>
<td>21</td>
<td>10</td>
<td>452</td>
</tr>
<tr>
<td>2</td>
<td>592</td>
<td>23</td>
<td>14</td>
<td>374</td>
</tr>
<tr>
<td>3</td>
<td>544</td>
<td>14</td>
<td>6</td>
<td>241</td>
</tr>
<tr>
<td>4</td>
<td>468</td>
<td>18</td>
<td>11</td>
<td>469</td>
</tr>
<tr>
<td>5</td>
<td>430</td>
<td>24</td>
<td>9</td>
<td>483</td>
</tr>
<tr>
<td>mean</td>
<td>515</td>
<td>20</td>
<td>10</td>
<td>404</td>
</tr>
</tbody>
</table>

Table 1. Lecture videos used.

<table>
<thead>
<tr>
<th>Lecture video</th>
<th>Word correct rate [%]</th>
<th>Word accuracy [%]</th>
<th>Out of vocabulary [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>59.5</td>
<td>46.3</td>
<td>3.9</td>
</tr>
<tr>
<td>2</td>
<td>55.7</td>
<td>44.6</td>
<td>2.2</td>
</tr>
<tr>
<td>3</td>
<td>53.9</td>
<td>42.3</td>
<td>3.7</td>
</tr>
<tr>
<td>4</td>
<td>41.3</td>
<td>30.3</td>
<td>3.6</td>
</tr>
<tr>
<td>5</td>
<td>54.5</td>
<td>41.7</td>
<td>3.8</td>
</tr>
<tr>
<td>mean</td>
<td>53.0</td>
<td>41.0</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Table 2. Speech recognition results for lecture videos.
Table 3. Keywords extracted by evaluators.

<table>
<thead>
<tr>
<th>Lecture video</th>
<th>OOV</th>
<th>Abstraction</th>
<th>The number of keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>9</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>10</td>
<td>152</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>8</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>17</td>
<td>115</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>8</td>
<td>84</td>
</tr>
<tr>
<td>Total</td>
<td>63</td>
<td>51</td>
<td>486</td>
</tr>
</tbody>
</table>

as association keywords by Japanese Wordnet. These results show that the keyword expansion is effective.

Figure 5 shows the result of investigating which relationship is important among hypernym and hyponym relationships in retrieval keyword expansion by Wordnet. In Figure 5, ‘All’ means the results using the broader concept (Y is a broader concept of X if every X is a kind of Y), the subordinate concept (Y is a subordinate concept of X if every Y is a kind of X), and entailed consequents of the retrieval keyword as association keywords by Japanese Wordnet. ‘Subordinate & Consequent’ means the results using the subordinate concept and entailed consequents of the retrieval keyword as association keywords. ‘Consequent’ means the results using entailed consequents. ‘Broader concept’ means the results using the broader concept. The result of expansion by using the subordinate concept was equivalent to the result of expansion by using the subordinate concept and entailed consequents of the retrieval keyword as association keywords. Moreover, the expansion by using the subordinate concept was superior to ‘All’.

3.3. Discussion

Table 3 shows OOV (out of vocabulary) and the number of unspoken words in subtopics among keywords extracted by the evaluators. In Table 3, ‘abstraction’ is a keyword that summarizes the content of the utterance, and expresses the event though not uttered directly such as “practice”, “explanation”, “outline”, “introduction”, and “summaries”. From Table 3, it was confirmed that measures against not only the unknown word but also the abstraction expression were important.

It was effective to expand the keyword by using the subordinate concept in Figure 5. This reason would be that retrieval unspoken keyword was converted into spoken keywords by the expansion to the subordinate concept when the retrieval keyword was abstracted.

In Table 3, OOV denotes the rate that the retrieval keyword doesn’t appear in the transcription text of the lecture videos. It was 13.0%. The keywords of 37.2% in the false alarm rate (FAR) did not appear in the ASR text. FAR has decreased to 26.9% by using the keyword expansion with Wordnet(subordinate concept). Moreover, FAR has decreased to 14.3% by using the keyword expansion with the dictionary.

3.4. Keyword expansion and subword retrieval

Figure 6 shows how using of the retrieval keyword expansion by the dictionary and the subword-based retrieval method for the retrieval keywords. ‘Subword’ means the results using the retrieval method of Figure 1 without keyword expansion, and the subword-based retrieval method for the retrieval keywords when the retrieval keyword consists of only an unknown word. ‘Subword only’ means the results using only subword-based retrieval method for the retrieval keywords.
Figure 6: Lecture subtopic retrieval results using keyword expansion by a dictionary and subword.

When keyword expansion ‘Dic’ and subword retrieval ‘Subword only’ are compared, ‘Subword’ is superior to ‘Dic’ when the number of key words consists of one, ‘Dic’ is superior to ‘Subword’ when the number of keywords is two or more. ‘Dic+Subword’ using the keyword expansion and subword retrieval is better than the individual use ‘Dic’ or ‘Subword’ to all the numbers of keywords.

Figure 7 shows how using of the retrieval keyword expansion by Wordnet and the subword together was effective. The results show that ‘Wordnet+Subword’ using the keyword expansion and subword retrieval is better than the individual use ‘Wordnet’ or ‘Subword only’ to all the numbers of keywords in queries.

4. Conclusions

The method of retrieving a subtopic of a lecture video was examined using the keyword expansion and subword retrieval. It was found that the expansion of the retrieval keyword by using the subordinate concept was effective. Moreover, the method using the keyword expansion and subword retrieval was better than the individual use.

In future works, we plan to investigate about the corpus of spoken Japanese lecture contents [15].

5. Acknowledgements

Part of this research was supported by Grant-in-Aid for Scientific Research (No. 19500845) from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

6. References