Analysis and Visualization of Index Words from Audio Transcripts of Instructional Videos

Alexander Haubold and John R. Kender
Department of Computer Science
Columbia University, New York
{ahaubold,jrk}@cs.columbia.edu
Overview

- Motivation
- Transcript Generation
- Target Corpus
- Filtering Index Phrases
- Filtering Word Pairs
- Visualization of Textual Information
- Future Investigations
Motivation

- Automatic Speech Recognition (ASR) Transcripts typically used for searching, categorization of video databases
- Lecture videos = dozens of contextually connected entities
- Typical course: 10 to 30 lectures (70 or 120 min), lecture: 5k – 14k words \( \approx \) 150k words
- Need indices across lectures, courses
- Extract and display structure of entire course using key words/phrases
... aum sell and it is its structure doesn’t provide a way to find something like a binary tree provides a way of looking for 27 and by treat it is given a pointer to the router the treehouse where it ought and emulate ...

Big Endian
Binary search
tree
...
Pointers

Interactive Visualization

Imperfect Transcript

Lecture Audio

Textbook Index
or Manual Index

Process Overview

find binary
find structure
find tree
binary tree
binary pointer
binary structure
tree pointer
tree structure
...

Index Words
and Phrases

Word Pairs
Transcript Generation

• ASR transcript from IBM® ViaVoice®
• Highly compressed lecture video
  – Poor audio quality
  – Word Error Rate of 75%
• Training: little (3%) overall improvement
  – Number of unique index phrases ≈ same
  – But: difference in identified index phrases 20%
  – Best to combine trained and untrained results
Target Corpus

- Lectures: rich in subject-specific terms
- Define:
  - *Theme phrases*: General tenor for contents of course
    - in many (> ¼) transcripts
  - *Topic phrase*: Highlight specific topics for lectures
    - in few (< ¼) transcripts
  - *Illustration phrases*: unique terms for examples
    - Hard to identify in highly imperfect transcripts
Filtering Index Phrases

- Structured approach
- Use corpus of expected phrases: index of course textbook
  - Capture key phrases of length 1-3
  - Rarely longer; index reflects likelihood
  - (1) Collapse indentation hierarchy
  - (2) Remove stop words in beginning & end of each line
  - (3) Stem

| amortized analysis | (3) (1,2,3) |
| accounting method of | (1,2) |
| aggregate analysis of | |
| ... | |
| call by value | |
Filtering Word Pairs

• Unstructured approach
  – Address: speech in lecture fragmented
• Use textbook index to filter keywords
  – Remove structure from index: use only words
  – Word pairs in transcript = index words separated by ≤ 10 words

<table>
<thead>
<tr>
<th>multiple instruction</th>
<th>call structural</th>
</tr>
</thead>
<tbody>
<tr>
<td>multiple operation</td>
<td>call hazard</td>
</tr>
<tr>
<td>multiple very</td>
<td>call instruction</td>
</tr>
<tr>
<td>multiple word</td>
<td>call compaction</td>
</tr>
<tr>
<td>multiple processor</td>
<td>call step</td>
</tr>
</tbody>
</table>
Results of Filtering

- Analysis of 273 transcripts (11 courses)
- Index Phrases
  - Unique per textbook indices: 253-4701
  - Unique per transcript: 8-98
  - Occurrence: 1 (35-50%), 5-50 (20-30%)
  - Unique per course: 40-347
- Word Pairs
  - \( \approx 10 \) times more than index phrases
  - Less meaningful for summaries
Interface: Parallels to a Camera

• 3 visualization techniques
• Share 3 freely variable parameters:
  – Zoom: specificity of phrases
    • Occurrence of phrase across transcript
    • Range: Topic-specific to entirely thematic
  – Focus: emphasis of phrases
    • Range: 1 – N (lowest – highest occurrence)
  – Contrast: length of phrases
    • Range: 1 – K (K usually 3)
Interface: Transcript Index Map

- Index phrases mapped to transcript
- Equivalent to textbook index
  - But: order by occurrence (highest near top)
  - Color coded (red→yellow = high→low occ.)
- Cross-reference terms among transcripts
  - Longer blobs for repeating phrases
- Greedy population of space near top
Demo

- Demo Transcript Index Map
Interface: Chapter Transcript Match

- Transcripts mapped to textbook chapters
- Rows=transcripts, Columns=chapters
- Match score based on occurrences of terms between transcript and chapter
- Performance:

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Overall best accuracy over all Zoom levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index Phrases + Word Pairs</td>
<td>0.7</td>
</tr>
<tr>
<td>Word Pairs</td>
<td>0.66</td>
</tr>
<tr>
<td>Word Pairs from G²</td>
<td>0.63</td>
</tr>
<tr>
<td>Index Phrases</td>
<td>0.5</td>
</tr>
</tbody>
</table>
• Demo *Chapter Transcript Match*
Interface: Transcript Similarity

- Cluster similar lectures (transcripts)
  - Dice distance of co-occurrence counts of selected index phrases
- Multidimensional scaling: |lectures| → 2D
• Demo Transcript Similarity
Interface: Transcript Index Map

- Zoom=1; “highly topic-specific”
### Interface: Chapter Transcript Match

- **Green** = correct, **Red** = incorrect, **Yellow** = other possibility

#### Transcript Analyzer

<table>
<thead>
<tr>
<th>Chapter Transcript Match</th>
<th>Transcript Index Map</th>
<th>Lecture Transcript Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>17</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>18</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>19</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>21</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>22</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>23</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>24</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>25</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>26</td>
<td>12</td>
<td>10</td>
</tr>
</tbody>
</table>

**Resize Columns to fit**
Interface: Transcript Similarity

- Course on “Visual Databases”
Future Investigations

• Incorporate tools into previously developed lecture browser
• User studies: usability of interface, ease of finding information
• Extend interface: provide index phrase layout on lecture level
• Extend indexing and visualization from intra-lecture to intra-course
Thank you!