SegChain:
Towards a Generic Automatic Video Segmentation Framework, based on Lexical Chains of Audio Transcriptions

Adrian-Gabriel CHIFU\textsuperscript{1}, Sébastien FOURNIER\textsuperscript{2}

LSIS UMR 7296 CNRS
Aix-Marseille Université
Marseille, France
\textsuperscript{1}adrian.chifu@lsis.org \textsuperscript{2}sebastien.fournier@lsis.org

June 14\textsuperscript{th}, 2016
Nîmes, France
Table of Contents

1. Story Segmentation Context
2. Lexical Chains
3. Segmentation Algorithm
4. Example on a Video
5. Preliminary Evaluation
6. Conclusion
1. Story Segmentation Context

- The need to retrieve information contained within videos
  - Automatic video segmentation
  - Characterize the semantically homogeneous segments

Domain specific (anchor, keywords, etc.) [Goyal et al., 2009]

News TV is only a small subcategory of videos on the web

The need for generic frameworks

Lexical chains to model topic changes

Hypothesis: within a video segment, there is a homogeneous distribution of the most frequent terms.
1. Story Segmentation Context

- The need to retrieve information contained within videos
  - Automatic video segmentation
  - Characterize the semantically homogeneous segments
- Numerous approaches on News TV
  - Domain specific (anchor, keywords, etc.) [Goyal et al., 2009]
  - News TV is only a small subcategory of videos on the web
  - The need for generic frameworks
1. Story Segmentation Context

- The need to retrieve information contained within videos
  - Automatic video segmentation
  - Characterize the semantically homogeneous segments
- Numerous approaches on News TV
  - Domain specific (anchor, keywords, etc.) [Goyal et al., 2009]
  - News TV is only a small subcategory of videos on the web
  - The need for generic frameworks
- Lexical chains to model topic changes
  - Hypothesis: within a video segment, there is a homogeneous distribution of the most frequent terms.
2. Lexical Chains

- **Lexical chains**
  - "Lexical cohesion ... over a succession of a number of nearby related words spanning a topical unit of the text" [Morris & Hirst, 1991]
  - Topic shifts are marked by the end of one or several lexical chains
2. Lexical Chains

- **Lexical chains**
  - "Lexical cohesion ... over a succession of a number of nearby related words spanning a topical unit of the text" [Morris & Hirst, 1991]
  - Topic shifts are marked by the end of one or several lexical chains

- **Hiatus**
  - The maximal distance between two term occurrences from the same lexical chain

\[
hiat(t_i) = \frac{\sum_{k=0}^{\lfloor \frac{|n0app(t_i)|-2}{2} \rfloor} \left[ n0app(t_i)_{k+1} - n0app(t_i)_k \right]}{|n0app(t_i)| - 1}, \quad i \in [0, N)
\]
2. Lexical Chains - Compactness

- **Chain compactness**
  - A term with dispersed chains is not representative for topics (similar to TF-IDF)

\[
\text{comp}(t_i) = \left\lceil \frac{n0app(t_i)|n0app(t_i)|^{-1} - n0app(t_i)_0}{|S|} \right\rceil \\
\times \frac{|n0app(t_i)| \times |\text{chain}(t_i)|}{\max_{j\in[0,N]} (\max (\text{lenCh}(t_j)))}
\]
3. Segmentation Algorithm

Algorithm 1 SegChain: The story segmentation method

Requires: A subtitle file in TRS or SRT format.
Ensures: Subtitle unit IDs (cuts) and frequent terms/segment.

1: if TRS file format then
2: convert file into SRT format
3: extract the raw text from the subtitle file
4: text normalization
5: $T \leftarrow$ the most frequent $N$ terms
6: build $S$
7: for all $t_i \in T$ do
8: compute $app(t_i, st_j)$, $hiat(t_i)$, $chain(t_i)$, $comp(t_i)$
9: compute $T'$
10: for all $St_j \in S$ do
11: compute $sim(st)$
12: compute $minima$
13: for all $St_j \in S$ do
14: compute $segment$
15: compute $front$
4. Example on a Video - Lexical Chains

Lexical chains of frequent terms along subtitle text
(Considered terms initially: 100) (no stemming)
4. Example on a Video - Lexical Chains

Lexical chains of frequent terms along subtitle text
(Considered terms initially: 100) (no stemming)
4. Example on a Video - Cosine Similarity

Cosine similarity for subtitle parts
(Considered terms initially: 75) (no stemming)

![Cosine similarity graph](image-url)
5. Preliminary Evaluation - Measures

**Precision**

\[ \text{Precision} = \frac{|\text{good\_front}|}{|\text{all\_detected\_front}|} \]
5. Preliminary Evaluation - Measures

**Precision**

\[ Precision = \frac{|good\_front|}{|all\_detected\_front|} \]

**Recall**

\[ Recall = \frac{|good\_front|}{|all\_true\_front|} \]
5. Preliminary Evaluation - Measures

**Precision**

\[
\text{Precision} = \frac{|\text{good\_front}|}{|\text{all\_detected\_front}|}
\]

**Recall**

\[
\text{Recall} = \frac{|\text{good\_front}|}{|\text{all\_true\_front}|}
\]

**F-measure**

\[
F - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]
5. Evaluation - Results

- French video

Table: Performance measures for TextTiling and SegChain, on the News TV video

<table>
<thead>
<tr>
<th>News TV video</th>
<th>TextTiling</th>
<th>SegChain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.1071</td>
<td>0.3333</td>
</tr>
<tr>
<td>Recall</td>
<td>0.4285</td>
<td>0.2143</td>
</tr>
<tr>
<td>$F$ – measure</td>
<td>0.1714</td>
<td>0.2609</td>
</tr>
</tbody>
</table>
5. Evaluation - Results

- **French video**

  Table: Performance measures for TextTiling and SegChain, on the News TV video

<table>
<thead>
<tr>
<th>News TV video</th>
<th>TextTiling</th>
<th>SegChain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.1071</td>
<td><strong>0.3333</strong></td>
</tr>
<tr>
<td>Recall</td>
<td><strong>0.4285</strong></td>
<td>0.2143</td>
</tr>
<tr>
<td>$F - measure$</td>
<td>0.1714</td>
<td><strong>0.2609</strong></td>
</tr>
</tbody>
</table>

- **English video**

  Table: Performance measures for TextTiling and SegChain, on the MOOC video

<table>
<thead>
<tr>
<th>MOOC video</th>
<th>TextTiling</th>
<th>SegChain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.0408</td>
<td><strong>0.1429</strong></td>
</tr>
<tr>
<td>Recall</td>
<td><strong>0.6667</strong></td>
<td>0.6667</td>
</tr>
<tr>
<td>$F - measure$</td>
<td>0.0769</td>
<td><strong>0.2353</strong></td>
</tr>
</tbody>
</table>
6. Conclusions

- **Conclusions**
  - SegChain: generic framework for video segmentation
  - Chain compactness measure
  - Evaluated on 2 videos (English & French) from 2 domains (MOOC & News TV)
6. Conclusions

- **Conclusions**
  - SegChain: generic framework for video segmentation
  - Chain compactness measure
  - Evaluated on 2 videos (English & French) from 2 domains (MOOC & News TV)

- **Future Work**
  - Building a MOOC corpus
  - SegChainW2V: semantic reasoning using word embedding for similarity
  - Temporal reasoning: Allen’s interval logic
Thank you for your attention!

- **Acknowledgements:** A*MIDEX project (nANR-11-IDEX-0001-02)
- **Follow us:** http://amidex.kalysee.com