Learning Weights in Manual Video Retrieval

Nov.18

Rong Yan
Carnegie Mellon University
Outline

• Framework and features
• Learning weights
• Experimental results
• Conclusion
Framework

Development Set

Anchor Detector

Face Detector

Text Search

Image Search

Negative PRF

Linear Fusion

Final Score

Carnegie Mellon
Features

• Text features
  • Automatic speech transcripts (ASR): sphinx, LIMSI
  • Video optical character recognition (VOCR)
  • Closed caption (CC)

• Image features
  • HSV color histogram: 150 bins
  • Gabor wavelet histogram: 108 bins
  • Canny edge histogram: 72 bins

• Face features
  • Presence of face
Modified Version of NPRF

• Negative pseudo-relevance feedback (NPRF)
  • Images farthest from query examples as the negative images
  • A support vector machine is learned to provide retrieval scores

• Modified sampling strategies
  • Negative samples should be dissimilar to each other
  • Maximal marginal irrelevance (MMIR) criterion
    - “Irrelevance” + “novelty”

\[
MMIR = \arg \min_{D_i \in T \setminus S} \left[ \lambda \text{Sim}_1(D_i, Q) + (1 - \lambda) \max_{D_j \in S} \text{Sim}_2(D_i, D_j) \right]
\]
Other semantic features

• Anchor detector
• Commercial detector
• Face recognition (in Person X queries)
Fusion

Development Set

Text Query

Image Query

Anchor Detector

Face Detector

Text Search

Image Search

Negative PRF

Linear Fusion

Final Score

Shots
Linear Fusion

• Weighted Borda fuse model
  • Generate rank lists from different features
  • Normalize output scores into rank-based scores
  • Linear combination

• Our focus
  • How to estimate the linear weights $w_i$
Learned by query types

- Prior information
  - *Finding Persons:* \((text \ 2, \ face \ 1, \ color \ 1, \ anchor \ 0)\).
  - Aircraft and animals: \((text \ 2, \ face \ -1, \ edge \ 1, \ anchor \ -1)\).
  - Others: \((text \ 2, \ face \ -1, \ color \ 1, \ anchor \ -1)\).
Learned by development set

- Supervised learning
  - Collect truth video shots in the development set using the Informedia clients in 15 minutes
  - Learning the weights using standard classifiers
Co-Retrieval

- Text retrieval results as feedback to learn weights
  - Inspired by multi-modal learning or co-training
- Algorithms
  - Grouping: Break the feature sets into two groups.
    - Text features vs. the other features;
  - Labeling: Label top K pseudo-positive data based on text
  - Learning: Use the pseudo-positive data to learn the linear weights
- Score combination: Combine the scores with the learned weights.
Descriptions of each run

- CMU-PX: FSD + Face recognition
- CMU-FSD: Tuned linear weighting with FSD truth
- CMU-400: Weighting from query types
- CMU-1000: Tuned linear weighting with FSD truth
- CMU-CoRet: CoRetrieval using top 100 shots as truth
- CMU1: as CMU-400 with 200 shots only
- CMUBase: text retrieval [ASR, CC + OCR]
- CMUBase1: text retrieval [ASR, CC]
# MAP Performance

<table>
<thead>
<tr>
<th>Runs</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMU-PX</td>
<td>0.218</td>
</tr>
<tr>
<td>CMU-FSD</td>
<td>0.207</td>
</tr>
<tr>
<td>CMU-400</td>
<td>0.198</td>
</tr>
<tr>
<td>CMU-1000</td>
<td>0.198</td>
</tr>
<tr>
<td>CMU-CoRet</td>
<td>0.195</td>
</tr>
<tr>
<td>CMU1</td>
<td>0.178</td>
</tr>
<tr>
<td>CMUBase</td>
<td>0.177</td>
</tr>
<tr>
<td>CMUBase1</td>
<td>0.155</td>
</tr>
</tbody>
</table>
Overall MAP for Manual Systems

"Manual" Search

System

MAP

CMU-1000
CMU-CoR
CMU-400
CMU-100
CMU-PX
CMU-FSD
CMU-BaseSYS
Results by Topics

Manual Retrieval Systems by Topic

MAP

Topic

CMU Base Text
CMU Co-Retrieval
CMU FSD+PersonX
CMU FSD Topic Train
CMU MM Base
CMU Top1000 txt
CMU Top400 txt
SYS 1
SYS 2
SYS 3
SYS 4
SYS 5
SYS 6
SYS 7
SYS 8
SYS 9
SYS 10
SYS 11
SYS 12
SYS 13
SYS 14
SYS 15
SYS 16
SYS 17
SYS 18
SYS 19
SYS 20
SYS 21
SYS 22
SYS 23
SYS 24
SYS 25
SYS 26
SYS 27
SYS 28
SYS 29

Carnegie Mellon
Conclusion
Estimate parameters

- Problem: How to estimate the parameters A & B
  - Manual Setting
  - Use rank-based score to approximate
    - Known as “Weighted Borda Fuse” in Meta-Search

\[
p(+|t = t_0) = Pr(t_0 < t < \infty) = 1 - D(t_0) = 1 - \frac{Rank(e)}{N}
\]