TRECVID 2010 Known-item Search by NUS

Xiangyu Chen, Jin Yuan, Liqiang Nie, Zhengjun Zha, Shuicheng Yan, Tat-Seng Chua

National University of Singapore, Singapore
Known-Item Search Task

- Given a text-only description of the video desired (Ground Truth Only One)

- Automatically return a list of up to 100 video IDs ranked by probability. (5 minutes)

- Interactively return the ID of the sought video and elapsed time to find it. (5 minutes)

**0022 QUERY:** Find the video of a man and woman getting dressed, a cat on window sill and another cat joining it, a wedding, two kittens and two babies
Motivations

- Efficient web service oriented video interactive search
- Efficient user interface (UI) for good interaction and efficient visualization
- New feedback algorithm based on both related samples and exclusive negative samples;
- Clustered shot-icons for fast previewing the main content of the videos.
VisionGo System

**User Interface**

- Maximize user’s annotation effort
- Video-Show: rich visual and audio content
- Clustering based Shot-Icons: Top-rank Icon + Expand Icon

**Auto Search**

- Multi-modality features fusion: Metadata, ASR, HLF and Youtube data
- Query Analysis

**Interactive Search**

- Related samples strategy
- Exclusive negative sample selection
- Fusion of two kinds of HLF
Efficient User Interface

Maximize user’s annotation effort

- **Video-Show**: show the detail and special visual and audio content

- **Clustered Shot-Icons**:
  Top-rank Icon + Expand Icon: represent the visual content of whole video
Efficient User Interface

- UI for good interaction and efficient visualization
- Maximize user’s annotation effort
Auto Search

Multi-modality features fusion
- Metadata is the most effective textual feature
- ASR plays a complementary role
- Tags of the crawled Youtube dataset

Query Analysis
- Query expansion by Youtube
- Morphological analysis between description of HLFs and KIS’s queries
Text query: Find the video of an Sega video game advertisement that shows tanks and futuristic walking weapons called Hounds.
Query Analysis

Query expansion by Youtube (two steps)

(a) Use the query to retrieve relevant video from Youtube and collect the tags/comments
(b) Extract terms from this collection (high mutual info.)

Morphological analysis

• HLF is necessary to query in terms of visual requirement
• Utilize WordNet to do selective expansion
• Match between feature descriptions of HLFs and KIS’s queries
Auto Search Performance

<table>
<thead>
<tr>
<th>Runs</th>
<th>Mean inverted rank</th>
<th>Mean elapsed time (mins)</th>
<th>Mean user satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run1 (Metadata+ Youtube)</td>
<td>0.215</td>
<td>0.021</td>
<td>6.0</td>
</tr>
<tr>
<td>Run2 (Metadata+HLF)</td>
<td>0.217</td>
<td>0.021</td>
<td>6.0</td>
</tr>
</tbody>
</table>

- Additional Tags data set is crawled from the Youtube website
- This dataset consists of 8,383 subsets of Youtube tags
- Each subset is downloaded corresponding to the title of each video

- Tags in Youtube are diverse as the words in metadata
- Need further denoise and extract key words in this dataset
Interactive Search

- Related Sample Strategy
- Exclusive Negative Samples Selection
- Fusion of Two Kinds of HLF
Related Sample Strategy

➢ Related Sample based Feedback

• Related sample refer to those video segments that are irrelevant to the query but relevant to some of the related concepts of the query. (Yuan el. CIVR10)
• New feedback strategy based on related shots of different videos

\[
f^t(x) = r^t \sum_{k=1}^{K} d^t_k f_k(x) + \frac{1}{t-1} \sum_{l=1}^{t-1} \beta^t_l \Delta f^l(x) + \Delta f^t(x)
\]

Shot query detector

Learn Video Detector by Fusion

\[
F^t(v_j) = \frac{1}{N_{v_j}} \sum_{p=1}^{N_{v_j}} f^t(s_p)
\]
Find the video: a man in orange outfit throwing an apple for a black dog with red collar to retrieve and the dog retrieves but eats the apple.

Transfer from video level to shot level.
Exclusive Negative Samples Selection

Exclusive Concept Subsets

\[ G_1 = \{ \text{airplane, infants, basketball, dancing, ... , hospital, maps, laboratory } \} \]
\[ G_2 = \{ \text{telephones, birds, chair, basketball, ... , flowers, golf, infants, maps} \} \]
\[ G_3 = \{ \text{laboratory, mountain, basketball, maps, ... , singing, kitchen, driver} \} \]

\[ \ldots \]
\[ G_{n-1} = \{ \text{golf, hospital, highway, infants, ... , laboratory, prisoner, stadium} \} \]
\[ G_n = \{ \text{boat_ship, cows, court, dancing, ... , computer_or_television_screen} \} \]

- If the selected related samples contain the concepts: “birds”, “mountain”, “highway”, then the exclusive negative set for the query is

\[ G_e = (G_2 \cup G_3 \cup G_{n-1}) \setminus \{ \text{“birds”, “mountain”, “highway”} \} \]

- Construction for exclusive concept sets:

  Robust Graph Mode Seeking by Graph Shift  \( (Liu\ H.\ and\ Yan\ S.\ ICML’10) \)
Fusion of Two Kinds of HLF

- Linear Fusion Detector Scores (130 concepts):
  Multi-label Propagation (Chen el. MM 2010) + CU-VIREO374 (Y.-G. Jiang el. 2008)

- Visual features:
  - 225-D blockwise color moments
  - 128-D wavelet texture
  - 75-D edge direction histogram

- Advantages:
  - Computation cost: about 32 hours
  - Learned concept scores are robust to noises
Interactive Search Performance

<table>
<thead>
<tr>
<th>Runs</th>
<th>Mean inverted rank</th>
<th>Mean elapsed time (mins)</th>
<th>Mean user satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run1 (Metadata+HLF)</td>
<td>0.628</td>
<td>2.799</td>
<td>5.75</td>
</tr>
<tr>
<td>Run2 (Youtube+HLF)</td>
<td>0.628</td>
<td>2.577</td>
<td>6.0</td>
</tr>
</tbody>
</table>

- Top 2 performance in all interactive search participants
- Validate proposed feedback scheme based on both related samples and exclusive negative samples
Interactive Search Performance

Find 15 out of 22 interactive topics
Demo of VisionGo

Interactive QUERYs:

• Find the video of a man and women getting dressed, a cat on window sill and another cat joining it, a wedding, two kittens and two babies
• Find the video of one girl in a pink T shirt and another in a blue T shirt doing an Easter skit with swirling lights in the background
• Find the video of 21 seconds of your time featuring orange, Japanese lanterns in the night
• Find the video of the cost of drugs, featuring a man in glasses at a kitchen table, a video of Bush, and a sign saying Canada
• Find the video of President Bush standing near sea vessels with Coast Guard members talking about his pride of the Coast Guard, immigration, and security issues.
• Find the video of a street that has a pedestrian crosswalk indicated with blue stripes. People are walking on the sidewalk and cars are driving on the street
Conclusions & Future Work

Contributions in this work

- Efficient UI in interactive video search
- Proposed feedback method based on both related samples and exclusive negative samples
- Clustered shot icons for fast previewing main content of the videos

Future work

- Extend the proposed novel feedback to real condition web services
- Develop more intuitive UI to enhance the user experience
Thank you!