Kobe University at TRECVID 2009 Search Task

Topic Retrieval based on Rough Set Theory and Partially Supervised Learning

Kimiaki Shirahama, Chieri Sugihara, Yuta Matsuoka and Kuniaki Uehara
System Overview

Difficulty of preparing indexing and retrieval models for all possible topics → Define a topic based on examples provided by a user

*Topic 289:* one or more people, each sitting in a chair, talking
Features

1. Grid-based color, edge and visual word histograms

![Grid-based color, edge and visual word histograms](image1)

2. Moving regions

\[ R = (x, y, \text{size}, h\_move, v\_move) \]

![Moving regions](image2)

3. # of faces with a certain size

One large-size face
Two small-size faces

One shot is represented by the Total 94 features!
Rough Set Theory

Large variation of features in the same topic
→ Extract **subsets** where positives can be correctly discriminated from all negatives

Topic 271: A view of one or more tall buildings …

Subsets are computed by boolean algebra of features and described by *decision rules*.

\[
\text{IF } \left( \begin{array}{c}
\text{Color hist. is similar to } \\
\text{Edge hist. is similar to }
\end{array} \right) \text{, THEN } \text{Positive}
\]
A great variety of shots can be negatives

*Topic 271: A view of one or more tall buildings (more than 4 stories) and the top story visible*

- **Too much dissimilar**
  → Many irrelevant features are included in decision rules

- **Neither similar nor dissimilar**
  Many relevant features are included in decision rules, e.g. long vertical edges, few edges in the upper part, etc.

- **Too much similar**
  → Many relevant features are ignored

How to select effective negatives for defining a topic?
Partially Supervised Learning

Build a classifier only from positives by selecting negatives from unlabeled examples

- Web document classification → Documents on the Web as unlabeled examples
- Our topic retrieval → Shots except for positives as unlabeled examples

**Similarity-based method** (Fung et al. TKDE 2006)
→ Effective in the case where only a small number of positives are available

- Reliable negative selection
- Clustering-based additional negative selection
Partially Supervised Learning

Build a classifier only from positives by selecting negatives from *unlabeled* examples

- Web document classification  →  Documents on the Web as unlabeled examples
- Our topic retrieval  →  Shots except for positives as unlabeled examples

**Similarity-based approach** (Fung et al. TKDE 2006)

→ Effective in the case where only a small number of positives are available

**Positives**

1. Reliable negative selection
2. Clustering-based additional negative selection

**How to calculate similarities in a high-dimensional feature space?**
Subspace Clustering

Due to many irrelevant features, we cannot appropriately calculate similarities → Find specific features to each example

**Subspace clustering** (*PROCLUS* proposed by C. Aggarwal et al. SIGMOD 99)

→ Group examples into *clusters in different subspaces of the high-dimensional space*

*Calculate similarities of an example to the other examples only by using the set of associated features!*
Submitted Runs

1. **M_A_N_cs24_kobe1_1**
   - Positives by manual and negatives by random

2. **M_A_N_cs24_kobe2_2**
   - Positives by manual and negatives by Partially Supervised Learning

3. **I_A_N_cs24_kobeS_3 (supplemental)**
   - Positives by manual and negatives by random
   - Positives and negatives interactively selected from each retrieval result

**Experimental purposes**
- Examine the effectiveness of rough set theory
- Examine the effectiveness of partially supervised learning
- Examine the Influence of positives and negatives on the performance
Example of Good Retrieval

*Topic 277:* A person talking behind a microphone

*Topic 285:* Printed, typed, or handwritten text, filling more than half of the frame area

*Topic 289:* One or more people, each sitting in a chair, talking

*Rough set theory can cover a large variation of features in the same topic!*
Comparison to Automatic Runs

NOTE: Only three runs have been submitted for the manually-assisted category.
Comparison to Interactive Runs

Difficulty of deriving an accurate conclusion for partially supervised learning

Why our runs are so bad?
### Additional Experiment

**Our assumption:** Features in submitted runs are ineffective

**Additional Experiment**

- Select 50 positives and 50 negatives from TRECVID 2008 test videos
- Use various combinations of features
- Features used in submitted runs:
  - Color, edge and visual word histograms,
  - Moving regions, # of faces with a certain size
- **Additional features:**
  - Grid-based color moment
  - Gabor texture
  - Concept detection scores (provided by MediaMill)
  - HOG
  - Camera work
- Retrieve shots of a topic in 200 of TRECVID 2009 test videos
Main reason for our bad runs

<table>
<thead>
<tr>
<th>Topic ID</th>
<th>271</th>
<th>272</th>
<th>287</th>
<th>291</th>
<th>292</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same features</td>
<td>14</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td><strong>Effective features</strong></td>
<td><strong>90</strong></td>
<td><strong>11</strong></td>
<td><strong>50</strong></td>
<td><strong>12</strong></td>
<td><strong>38</strong></td>
</tr>
<tr>
<td><em>Estimated best values</em></td>
<td>70</td>
<td>22</td>
<td>86</td>
<td>22</td>
<td>10</td>
</tr>
<tr>
<td>Best values in TRECVID ’09</td>
<td>209</td>
<td>66</td>
<td>257</td>
<td>66</td>
<td>30</td>
</tr>
</tbody>
</table>

*Estimated best values* 70 22 86 22 10
Best values in TRECVID ‘09 209 66 257 66 30

**Using ineffective features is the main reason for our bad runs!**

- Promising performance when effective features can be selected
- Effectiveness of camera work feature
Using ineffective features is the main reason for our bad runs!

- Promising performance when effective features can be selected
- Effectiveness of camera work feature
What is an Effective Feature?

<table>
<thead>
<tr>
<th>Topic ID</th>
<th>271</th>
<th>272</th>
<th>287</th>
<th>291</th>
<th>292</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original result</td>
<td>72</td>
<td>8</td>
<td>34</td>
<td>9</td>
<td>24</td>
</tr>
<tr>
<td>Original features</td>
<td>Concept</td>
<td>Concept + Color mom.</td>
<td>Concept</td>
<td>Concept</td>
<td>Camera work + # of faces</td>
</tr>
<tr>
<td>Best result</td>
<td>90</td>
<td>11</td>
<td>50</td>
<td>12</td>
<td>38</td>
</tr>
<tr>
<td>Effective features</td>
<td>Color hist.</td>
<td>Color hist.</td>
<td>Camera work</td>
<td>Gabor tex.</td>
<td>Concept</td>
</tr>
<tr>
<td>Worst result</td>
<td>76</td>
<td>2</td>
<td>16</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>All features</td>
<td>66</td>
<td>7</td>
<td>19</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Posteriori Comb.</td>
<td>80</td>
<td>4</td>
<td>36</td>
<td>4</td>
<td>37</td>
</tr>
</tbody>
</table>

**Features**
- Color hist. + Edge hist. + Color mom. + Camera work
- Color hist. + Gabor tex.
- Concept + Color mom.

**Rather than many features, using two or three features leads to the best performance!**

Neither visual words nor HOG are effective features.
How Retrieved Shots Change Depending on Features?

<table>
<thead>
<tr>
<th>Topic ID</th>
<th>271</th>
<th>272</th>
<th>292</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original result</td>
<td>72</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>Original Feature</td>
<td>Concept</td>
<td>Concept</td>
<td>Camera work + # of faces</td>
</tr>
<tr>
<td>Overlapping shots</td>
<td>66</td>
<td>61</td>
<td>28</td>
</tr>
<tr>
<td>Removed shots</td>
<td>6</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Added shots</td>
<td>24</td>
<td>15</td>
<td>16</td>
</tr>
</tbody>
</table>

**NOTE:** Similar results are obtained for Topic 287 and 291

**++** Effective features preserve many relevant shots retrieved by original features, and add more relevant shots.

**--** Ineffective features remove many relevant shots retrieved by original features.
How Decision Rules Change Depending on Features?

<table>
<thead>
<tr>
<th>Topic 271: Tall building</th>
<th>Building</th>
<th>Sky</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept (Original)</td>
<td>357</td>
<td>210</td>
<td>385</td>
</tr>
<tr>
<td>Concept + color hist.</td>
<td>361</td>
<td>204</td>
<td>342</td>
</tr>
<tr>
<td>Concept + Gabor tex.</td>
<td><strong>241</strong></td>
<td><strong>152</strong></td>
<td><strong>327</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topic 287: People, table and computer</th>
<th>Face</th>
<th>Office</th>
<th>Computer or Television</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept (Original)</td>
<td>177</td>
<td>284</td>
<td>235</td>
</tr>
<tr>
<td>Concept + Camera work</td>
<td>138</td>
<td>355</td>
<td>174</td>
</tr>
<tr>
<td>Concept + Edge hist.</td>
<td><strong>77</strong></td>
<td>303</td>
<td><strong>86</strong></td>
</tr>
</tbody>
</table>

++ Effective features preserve most of useful decision rules
-- Ineffective features substitute useful decision rules with inaccurate ones
How to Select Negatives?

<table>
<thead>
<tr>
<th>Topic ID</th>
<th>271</th>
<th>272</th>
<th>287</th>
<th>291</th>
<th>292</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>80 (+8)</td>
<td>3 (-5)</td>
<td>58 (+24)</td>
<td>12 (+3)</td>
<td>33 (+9)</td>
</tr>
<tr>
<td>Features</td>
<td>Concept</td>
<td>Concept + Color mom.</td>
<td>Concept</td>
<td>Concept</td>
<td>Camera work + # of faces</td>
</tr>
<tr>
<td>Best result</td>
<td>92 (+2)</td>
<td>8 (-3)</td>
<td>56 (+6)</td>
<td>15 (+3)</td>
<td>36 (-2)</td>
</tr>
</tbody>
</table>

| Topic 287: one or more people, each at a table or desk with computer visible |

Random
- Many edges in the upper part
- Many shots where a person appears

Partially supervised learning
- Few edges in the upper part
- Small number of shots where a person appears

Near miss negatives are not useful for defining a topic in videos!
Conclusion and Future Works

Conclusion:

*Example-based topic retrieval system*

- **Rough set theory** for covering a large variation of features in a topic
  - Relevant shots containing significantly different features can be retrieved.
- **Partially supervised learning** for negative example selection
  - Selected negatives are more useful than negatives selected by random

**But,** much more improvement is needed for a satisfactory retrieval!

Future works:

- Learning a **similarity measure** which is closely associated with human perception, by using training image pairs labeled as “similar” or “dissimilar”
- Constructing an **event ontology** in order to retrieve an event by considering its relation to the other events
- Developing a **browser** which enables users to easily collect a sufficient number of positives and negatives.
Thank you!