FIU-UM at TRECVID 2017: Rectified Linear Score Normalization and Weighted Integration for Ad-hoc Video Search

Submission Details

- Class: M (Manually-assisted runs)
- Training type: D (IACC & non-IACC non-TRECVID data)
- Team ID: FIU-UM (Florida International University - University of Miami)
- Year: 2017
Outline

- Introduction
- The Proposed Framework
- Experimental Results
- Conclusion and Future Work
Introduction
TRECVID 2017

- Year 2015: Semantic indexing (SIN)
- Year 2016: Ad-hoc video search (AVS)
- Year 2017: Same training and testing datasets, different topics

- Test collection: IACC.3
- 346 concepts
- 30 Ad-hoc queries
- Submit a maximum of 1k possible shots from the test collection for each query
The Proposed Framework
CNN Feature Extraction

- Last Pooling Layer
- Feature: ImageNet-1000
Classification

- Support Vector Machine (SVM)
- Linear kernels
- Positive weight / Negative weight: 1:1
Rectified Linear Score Normalization

- How to eliminate the effect of “bad” scores of a concept in an Ad-hoc query before the score fusion

- Two thresholds:
  - threshold_high
  - threshold_low
Algorithm 1 The proposed rectified linear score normalization algorithm.

for all scores do
  if $score \geq threshold_{high}$ then
    $score_{normed} = 1$;
  else if $score \leq threshold_{low}$ then
    $score_{normed} = 0$;
  else
    $score_{normed} = \frac{score}{threshold_{high}}$.
  end if
end for
Rectified Linear Score Normalization
Query Formulation and Score Combination

- More concepts:
  - A pretrained ImageNet model: ImageNet1000

- Score fusion:
  - Weighted geometric mean

\[ score_{query} = \prod_{i}^{N} (score_{concept_i})^{weight_i} \]
Experimental Results
Model training: using TRECVID 2010-2012 training videos as the training data

Model evaluation: using TRECVID 2013-2015 training videos as the testing data to evaluate the framework and tune the parameters of the models

Model testing: using TRECVID 2010-2015 training videos as the TRECVID 2017 training data, and TRECVID 2017 testing videos as the testing data to generate the ranking results for the submission
Mean extended inferred average precision (mean xinfAP)
allows the sampling density to vary so that it can be 100% in the top strata. This is the most important one for average precision

As in the past years, other detailed measures based on recall and precision are generated and given by the sample eval software provided by the TRECVID team
Four Runs Submitted

- 1: CNN features + Linear SVM
- 2: CNN features + Linear SVM + Scores from other groups
- 3: CNN features + Linear SVM + Rectified Linear Score Normalization
- 4: CNN features + Linear SVM + Scores from other groups + Rectified Linear Score
## Performance

<table>
<thead>
<tr>
<th>Framework</th>
<th># of inferred true shots returned</th>
<th>Mean xinfAP values</th>
</tr>
</thead>
<tbody>
<tr>
<td>M_D_FIU_UU.17.1</td>
<td>3608</td>
<td>0.088</td>
</tr>
<tr>
<td>M_D_FIU_UU.17.2</td>
<td>4731</td>
<td>0.147</td>
</tr>
<tr>
<td>M_D_FIU_UU.17.3</td>
<td>4125</td>
<td>0.102</td>
</tr>
<tr>
<td>M_D_FIU_UU.17.4</td>
<td>4623</td>
<td>0.145</td>
</tr>
</tbody>
</table>
Performance

Run1

Run3

Inferred avg. precision

Run score (dot) versus median (---) versus best (box) by concept
Performance

Run2

Run score (dot) versus median (---) versus best (box) by concept

Run4
Conclusion and Future Work
Conclusion and Future Work

- In our framework, only global features are currently utilized => the object-level features can also be explored by R-CNN (Regional CNN)
- Non-linear SVM classifiers need to be adopted to address the data imbalance issue
- More advanced CNN structures can be integrated and scores from them can be fused
- Temporal correlations can be considered to reach a better performance
- More training data should be collected by a general purpose search engine like Google using the query definition to further improve the retrieval accuracy
THANKS!

Any questions?