Merging Results from Multiple Sources in Video Classification

Wei-Hao Lin
Language Technologies Institute
School of Computer Science
Carnegie Mellon University
whlin@cs.cmu.edu

Multiple-Source Video Classification

Find video segments containing “Madeleine Albright”

CNN
ABC
C-SPAN

11-743 Advanced Information Retrieval Seminar and Lab (Mid-Term Project Presentation)
Problem: Merging Results

Motivations

- Collections may not contain the same type document.
  - Music + Image + Speech Transcript + …
- Full data set may take long time to train.
  - Especially for classifiers that time complexity is not linear
- There may exist collection-specific characteristics that are useful for video classification.
Transform the score from individual source (local score, $L(d)$) to the score that is comparable to scores from other sources (global score, $G(d)$).

- $G(d) = L(d)$
- $G(d) = T(L(d))$

**Score-based Local Functions**

- **Raw Score**
  - $G(d) = L(d)$

- **Linear Scaling (Normalized Score)**
  - $G(d) = \frac{[L(d) - \min L(d)]}{[\max L(d) - \min L(d)]}$

- **Posterior Probability**
  - $G(d) = \Pr(\text{Relevant} | d)$
  - Logistic regression is used to transformed SVM output to probability.
**Rank-based Local Functions**

- **Round-Robin**
  - \( G(d) = -\text{Rank}(d) \)
- **Borda Vote**
  - \( G(d) = N - \text{Rank}(d) \)
- **Normalized Borda Vote**
  - \( G(d) = \frac{N - \text{Rank}(d)}{N} \)

---

**Local-to-Global Mapping Function**

**A**
- Train on T/T' to generate local scores
- Train T+T' to learn mapping function

**B**
- Train on T/T' to generate local scores of V and V'
- Train T+T'+V+V' to learn mapping function
**Experiment Testbed**

- TREC 2003 Video track Feature Development Set
  - Testbed: Broadcast News
  - Source 1: ABC, ~26 hours, 25630 shots
  - Source 2: CNN, ~28 hours, 21696 shots
  - Source 3: C-SPAN, ~6 hours, 776 shots

**Metric: Mean Average Precision**

- Example: \((\frac{1}{2} + \frac{2}{5}) / 2 = 0.45\)
- Between 0 and 1; the higher, the better
- Focus on order, and emphasize top-ranked items
- Random baseline of \(m\) relevant items in the list \(n\)
  - Approximation: randomly assign \(m\) positions
  - Analytic solution: calculate \(E[\text{AP}(m, n)]\)
17 Feature Detection Tasks

- Outdoors
- News subject face
- People
- Building
- Road
- Vegetation
- Animal
- Female Speech
- Car/Truck/Bus
- Aircraft
- News subject monologue
- Now-studio setting
- Sporting event
- Weather news
- Zoom in
- Physical violence
- Person X (Madeleine Albright)

Text Classification

- Document: Closed Captions or Speech Transcripts in the shot
- Weighting scheme: tfc
- Stopword removal
- Porter stemming
- Classifier: C-SVM with linear kernel
11-743 Advanced Information Retrieval Seminar and Lab (Mid-Term Project Presentation)

Rare Class

Feature 3 - People

11-743 Advanced Information Retrieval Seminar and Lab (Mid-Term Project Presentation)
Merging Results - Feature 3 People

10-fold CV MAP

ABC+CNN+CS-PAN  raw score  linear scaling  round robin  probability  Random Baseline

Feature 14 - Weather News

10-fold CV MAP

ABC  CNN

Random Baseline
Preliminary Results

- Score-based or Rank-based methods fail dramatically when the assumption is not valid.
- If the classifiers are the same, raw score appears to be more comparable.
Next Steps

- Implement Local-to-Global mapping function.
- Complete full set of classification tasks.
- Conduct evaluation using other modal features, including Image, Audio, Face.