Concept Detection Based on LDA-SVM

Sheng Tang (唐胜), Jin-Tao Li (李锦涛), Ming Li (李明)
Cheng Xie (谢呈), Yi-Zhi Liu (刘毅志), Kun Tao (陶焜), Shao-Xi Xu (徐邵稀)
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, 100080
Email: ts@ict.ac.cn Tel: 8610-62600617

For TRECVID 2008 concept detection task, we focus on:

- To improve the training efficiency and explore the knowledge between concepts or hidden sub-domains, we propose a novel method based on Latent Dirichlet Allocation: LDA-based Multiple-SVM (LDA-SVM);
- Early fusion of texture, edge and color features TECM: TF*IDF weights based on SIFT features + Edge Histogram+ Color Moments;
- Introduction of Pseudo Relevance Feedback (PRF) into our concept detection system for the purpose of making re-trained models more adaptive to the test data.
1 LDA-SVM

1.1 Flowchart of LDA-SVM

1.2 Topic-simplex Representation Vector (TRV)

1.3 Our Novelties

- **Sample’s separability-keeping strategy during training**
  Unlike multi-bag SVM, we only use positive samples in current topic for the sake of retaining sample’s separability, instead of all positive samples among the whole training set, and ignore the topics with too few positive samples.

- **TRV-weight-based fusion strategy during testing**
  While testing a keyframe for a given concept, we adopt TRV as the weight vector, instead of equal weighting strategy, to combine the SVM outputs of topic-models.
2 System overview

2.1 Early Fusion

Early fusion of texture, edge and color features TECM (890 dims), abbreviation of the combined:
- TF*IDF weights based on SIFT features (345 dims)
- Edge Histogram (320 dims)
- Color Moments (225 dims).

2.2 Novel LDA-SVM Detection Method

- LDA clustering
  After quantization of the TF*IDF weights, we use Latent Dirichlet Allocation to cluster all the keyframes into 20 topics according to the maximum element of the TRVoF each keyframe.
- SVM Training
  ■Sample’s separability-keeping strategy
  For all the 20 concepts, we get 344 models after removing 56 topics with no more than 1 positive sample.
- SVM Test
  ■TRV-weight-based fusion strategy

2.3 Pseudo Relevance Feedback (PRF)

Unlike existing PRF techniques in text and video retrieval, we propose two preliminary strategies to explore the visual features of positive training samples to improve the quality of pseudo positive samples:
- Similarity-based method
  Select pseudo positive samples by calculating the feature similarities between top-retrieval examples with positive training samples after every retrieval process.
- Detector-based method
  Select pseudo positive samples through the overall evaluation of positions among the ranked lists from several detectors.

2.4 Object-based features

Object-based features: we train models with object-based TF*IDF features within labeled rectangles for positive training samples. But our result is not good due to unavailability of such object-based features of test samples.
3 Annotation & Experiments

3.1 Annotation of training data

![Annotation Interface](image)

Fig 3 The interface for our annotation

In order to encourage researchers to propose methods extracting features based on object rather than the whole frame, we divided the 20 concepts into two groups:

1. Object-related concepts
2. Scene-related concepts

3.2 InfAP of our runs

<table>
<thead>
<tr>
<th>HLF run</th>
<th>InfAP</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_ICT_1</td>
<td>0.048</td>
<td>Visual Baseline</td>
</tr>
<tr>
<td>A_ICT_2</td>
<td>0.038</td>
<td>LocalizationClassifier</td>
</tr>
<tr>
<td>A_ICT_3</td>
<td>0.065</td>
<td>TECM_LDA_SVM</td>
</tr>
<tr>
<td>A_ICT_4</td>
<td>0.037</td>
<td>TECM_LDA_SVM_PRF</td>
</tr>
<tr>
<td>A_ICT_5</td>
<td>0.076</td>
<td>TECM_LDA_SVM+Baseline</td>
</tr>
<tr>
<td>A_ICT_6</td>
<td>0.078</td>
<td>Fusion All</td>
</tr>
</tbody>
</table>

3.3 Result Analysis

- Effective: 35.4% improvement (run3 via run1)
- Efficient:
  - Topic size is greatly smaller
  - Samples in each topic are of higher separability
  - SVM training is very efficient, only about 20 minutes for all the 344 models on our cluster server (dualcore 1.8ghz *15)
  - Employing all samples in each topic for cross-validation becomes very practicable (about 12 hours for all 344 model on our cluster server).

3.4 Conclusion

1. Early fusion TECM, clustering via LDA, sample’s separability-keeping strategy, and TRV-weight-based fusion strategy together contribute to the high efficiency and effectiveness of our proposed method.
2. Determination method of hidden topic number should be carefully studied for further improvement.
3. PRF method is not stable since the introduction of pseudo positive samples may ruin the separability of topic samples.
4. More frames per shot should be used for test data.
5. Should combine LIG annotation to remove false annotations.