Inputs to TV2011 planning

17. November 2010

CCD

- Only first year for Internet video, resources in place, so continue task as is for another year
- Make any data problems (e.g., audio out of sync with video) known along with fixes if any
- Try to add new transformations
- Multiple ref segments in a query?
- Provide optimal results if only one threshold (for all transformations) can be set.
- Evaluate audio alone and video alone in addition to a+v

SIN

- Same or similar task.
- Same type of data.
- Similar volume of data? Or still more?
- A third (Large scale, ~1000) set of concepts?
- Subtasks, e.g. persons, events, actions, locations, genres ...?
- Other classes of concepts? Emotions?
- Multiple levels of relevance for positive samples?
- Or ranking of positive samples?
- Encourage and provide infrastructure for sharing contributed elements: low-level features, detection scores, ...
- Possibility to submit unpooled runs to encourage the evaluation of the effect of many parameters.
- Derived measure: GMAP to better recognize work on difficult concepts?

KIS

- Continue the task
- Another 200 hrs of video and 300 topics available
- Is the task too hard, yet its not unrealistic
- Could topic formation benefit from memory science input, recalling the memorable characteristics of KI some time (weeks?) after viewing ?
- Do the 67 'never found' videos have no metadata or very little metadata ? Need to do a failure analysis of what made some topics irretrievable, or more difficult than others.
- How do we know there is just one true groundtruth … can we post-process the submitted results to see if there are other non-KI videos which appear regularly indicating they are close, or near-duplicates of the KI
- Metadata exploitation was key (e.g. I2R got a lot from linguistic analysis) but has metadata exploitation peaked ? Metadata links ?
- CMU got a lot from query-type classification leading to different fusion types - should we unify query types ?
- Content-based techniques didn't help … is this "didn't help YET" ?

INS

- Continue the task, perhaps as pilot again
- Be strict about the notion of “instance” (identity) to delimit the task from others
- Use more topics if possible.
- Avoid topics with only a few instances in the test data
- Topic types ok
- No one mined video for variety of target examples!
- Maybe drop some versions of the masked target from the topic
- Avoid extremely small target examples in the topics
- Look for series/movies in the IACC data with recurring people and places (we have the data and shots)
- Look for pubicly available series/movies from other Internet sources
- Try queries for products – if video appropriate
- Measures ok – precision, recall, AP - but perhaps emphasize precision
- Any HD video sources available?

SED

- Continue down the same path
  - Re-use the evaluation data set
  - Share annotations

MED

- Reactions to the pilot
  - Concern over event richness within the test corpus, needs to be lower
  - Observation that static detectors worked well: need events with strong temporal requirements
  - Are cost measures warranted with no temporal component to evaluate?
- Reactions to the MED ‘11 proposal
  - 15 events better than 10 events
  - Make MED ‘10 data available ASAP (before Mar 1, 2011)
  - Share annotations and detectors