Crowdsourcing for Affective Annotation of Video: Development of a Viewer-reported Boredom Corpus

Mohammad Soleymani\textsuperscript{1}, Martha Larson\textsuperscript{2}

\textsuperscript{1}Computer Vision and Multimedia Lab.
University of Geneva, Switzerland

\textsuperscript{2}Multimedia Information Retrieval Lab.
Delft University of Technology, the Netherlands
Outline

• Background and motivation
• MediaEval 2010 affect task
• Affective computing corpora
• Two step crowdsourcing scheme
• Analysis of annotations
• Best practices
MediaEval Affect task 2010

- **Use Scenario:** User would like interesting content to be recommended
- **Task:** Rank videos with respect to user perceived boredom
- **Data:** SPUG video series from blip.tv
- **Groundtruth:** Generated by human assessors
Which one is boring?
Previous work in corpus development

• Psychological datasets (conventional)
  – Philippot, 1993
  – Rottenberg et al, 2007

• Our previous work (online)
  – Online annotations from more than 40 participants
  – 1300 annotations on 155 videos
    (Soleymani et al, ACII 2009)
Motivation

• Limitations of the previous corpora
  – Licensing and copyright
  – Limited resources
  – The whole collection annotated with as many as possible

• What is added with crowdsourcing
  – Large number
  – Diversity
  – Target population
Amazon Mechanical Turk

- Crowdsourcing platform that makes possible micro outsourcing of tasks
- Micro-tasks called Human Intelligence Tasks (HITs)
- HITs are carried out by MTurk workers (turkers)
- Typically used for tasks that lend themselves well to piecemeal work (multiple people make small contributions)
- Requesters can assign qualifications to turkers
High commitment crowdsourcing

• A single turker is needed to carry out a large set of HITs
• Different from typical piecemeal tasks
• 125 videos had to be annotated
• Two step approach
  – Step 1: qualification and personal information
  – Step 2: Carrying out the series by qualified turkers
First step HITs and qualifications

• Only turkers with HIT acceptance rate >95%
• Qualification based on the performance on the first step HIT
• Assigning the qualification and inviting for the main HIT
Second step HIT

Qualifications Required: PetamediaVideoAnnotationQualification has been granted

HIT Preview

Which word best describes the emotion you felt while watching the clip?
- please choose -
Describe the documentary in one sentence

How boring was the video? (1 = very boring and 9 = not at all boring)
1 2 3 4 5 6 7 8 9
boring dislike

How much did you like the video? (1 = I hated it and 9 = I liked it a lot)
1 2 3 4 5 6 7 8 9
not boring like

How long was the video clip? (write your estimation, for example, three minutes and half will be written 3:30, write 3 in the minutes box and 30 in the seconds box)

7/23/2010 CSE workshop, SIGIR 2010
Second step HIT

• Target information
  – Self reported boredom score
  – Self reported like/dislike rating
  – Time perception

• Context information
  – Time of day
  – Mood word question

• Validation Question
  – Description
Analysis of annotations

• Pilot HIT 169 workers
  – 88% watch online videos on internet
  – Gender: 105 male 62 female and 2 unknown!
  – Age: Mean = 30.5 STD = 12.4

• 47% of the turkers in the first step carried out the single HIT completely and earned the qualification

• 40% of the qualified turkers skipped parts of the videos
Best practices

• The step approach worked well for our high commitment task

• For high commitment tasks, five times as many workers are needed to be invited for the first step

• Establishing trust
  – Interacting with workers
  – Granting bonuses
  – Accepting HITs as quickly as possible