

Affective Impact of Movies: Task Overview and Results

Mats Sjöberg, Yoann Baveye, Hanli Wang, Vu Lam Quang,
Bogdan Ionescu, Emmanuel Dellandréa, Markus Schedl,
Claire-Hélène Demarty, Liming Chen

MediaEval 2015 Workshop, Wurzen, September 14-15



JOHANNES KEPLER
UNIVERSITY LINZ | JKU



Affect Task in 2015

- Fifth year of “Affect Task” with evolving topic
- Use scenario: automatic tools to help users find videos that fit their particular mood, age or preferences
- Two subtasks to address this:
 - **Induced affect detection** (emotional impact)
 - **Violence detection**
- New Creative Commons-licensed data set this year
- Transition year ...

Task definition

- **Induced affect detection** or emotional impact:
 - Valence: *negative, neutral, and positive*
 - Arousal: *calm, neutral, and active*
 - Evaluated by *accuracy* – classification task
- **Violence detection:**
 - Content that “*one would not let an 8 years old child see because they contain physical violence*”
 - Evaluated by *average precision* – retrieval task

Dataset

- 10,900 short (8-12 secs) video clips
- Extracted from 199 professional and amateur movies
- Creative Commons-licensed → legal to share
- Basic set of features
- Based on LIRIS-ACCEDE <http://liris-accede.ec-lyon.fr/>

	LIRIS-ACCEDE		EXTENSION	TOTAL
Movies	100	60	39	199
Clips	6,144	3,656	1,100	10,900
	<i>– dev. set –</i>	<i>– test set –</i>		

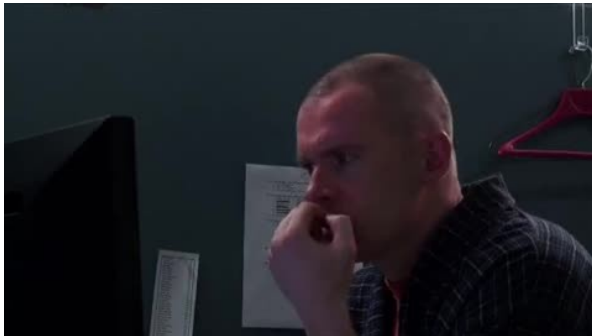
Ground truth annotations: affect

- LIRIS-ACCEDE: valence-arousal scores included
- Crowd source *quicksort* to get full ranking:
 - Select pivot randomly
 - Divide dataset into two by pair-wise ranking
 - Repeat for subsets
- Transformed into absolute affective scores by experts and interpolation: $[-1, 1]$

Ground truth annotations: affect, cont.

- For MediaEval valence/arousal split into three classes from absolute scores at limits -0.15 and 0.15
- MediaEval extension annotated with simplified scheme: manually selected two pivot

negative

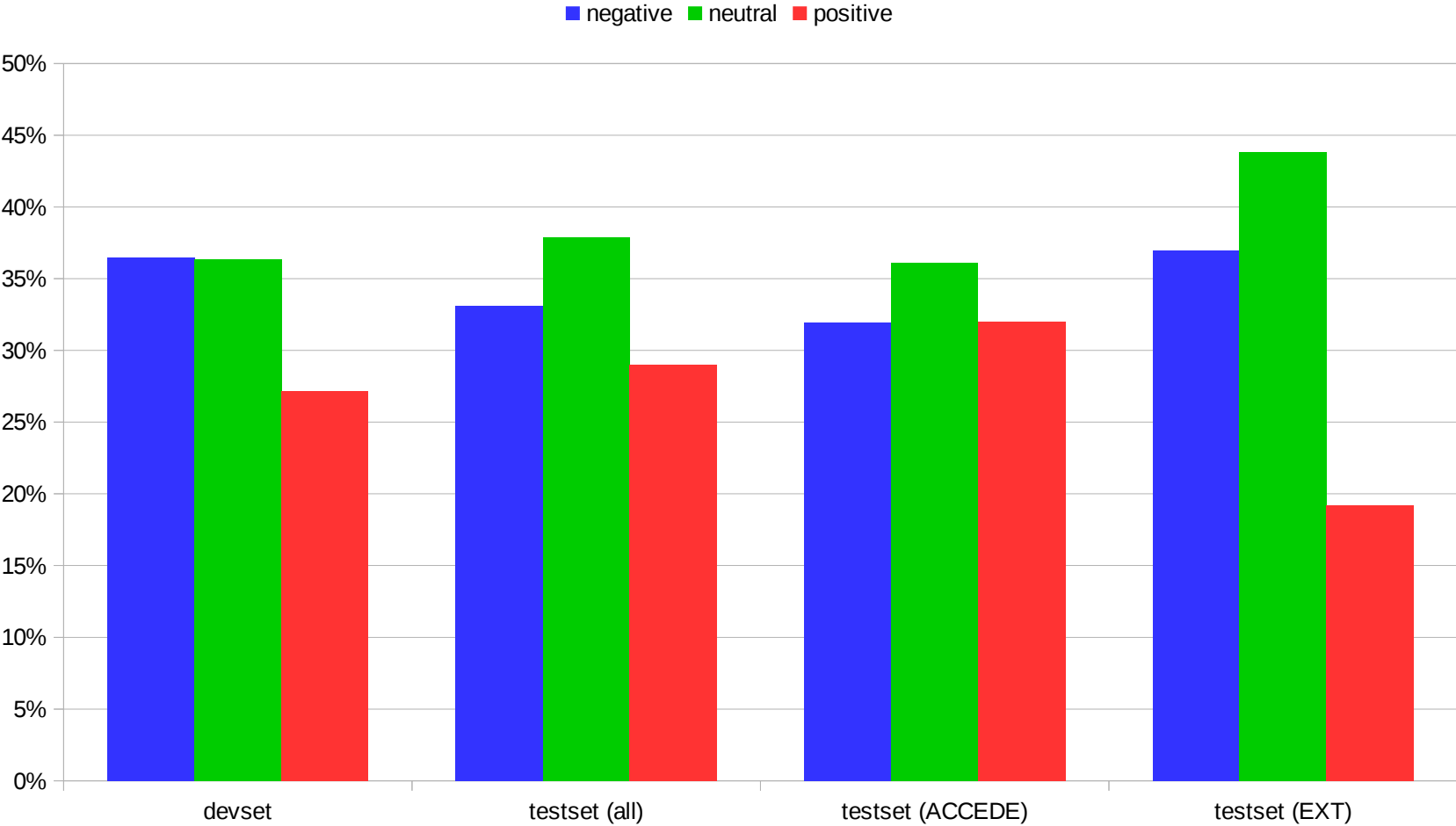


neutral

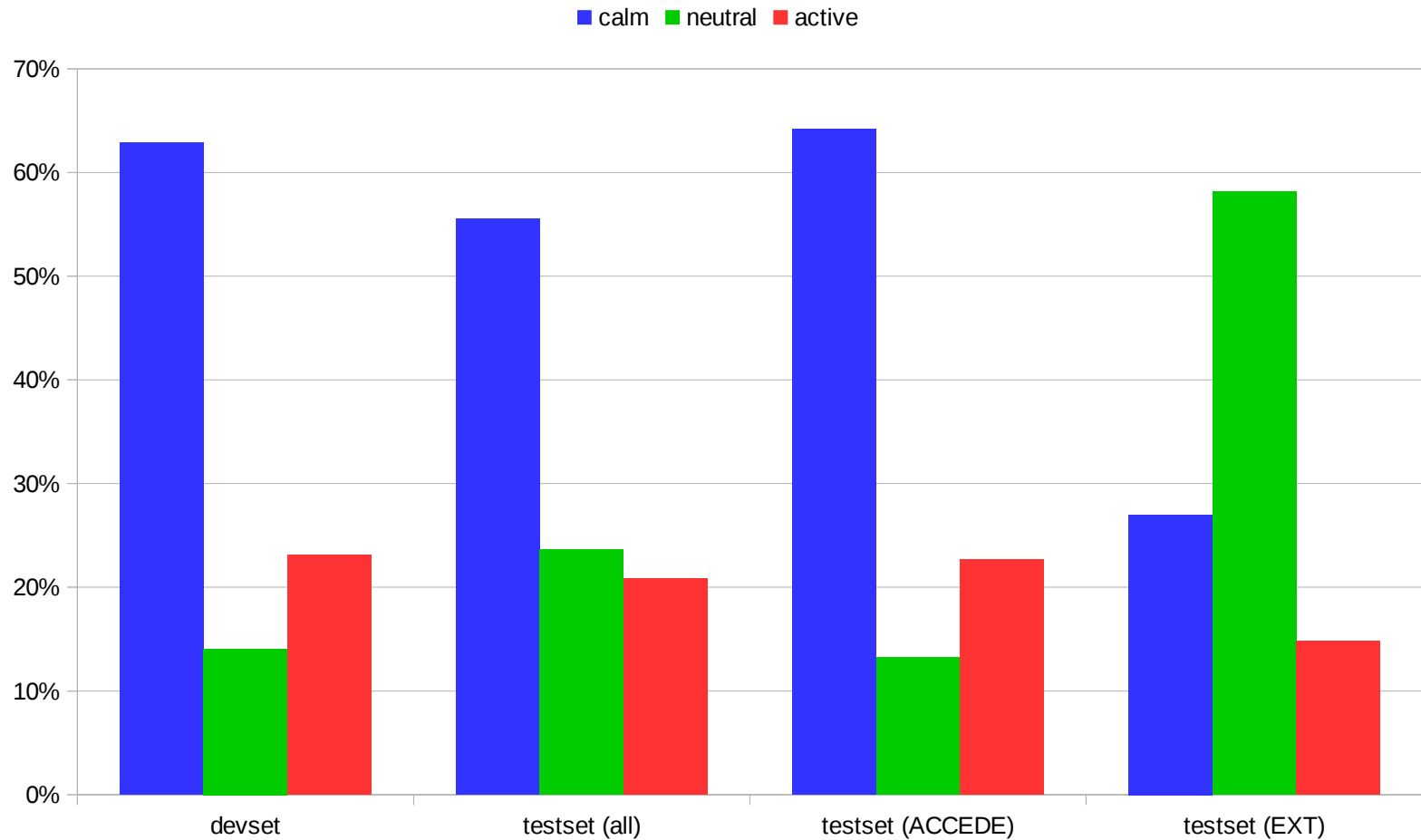


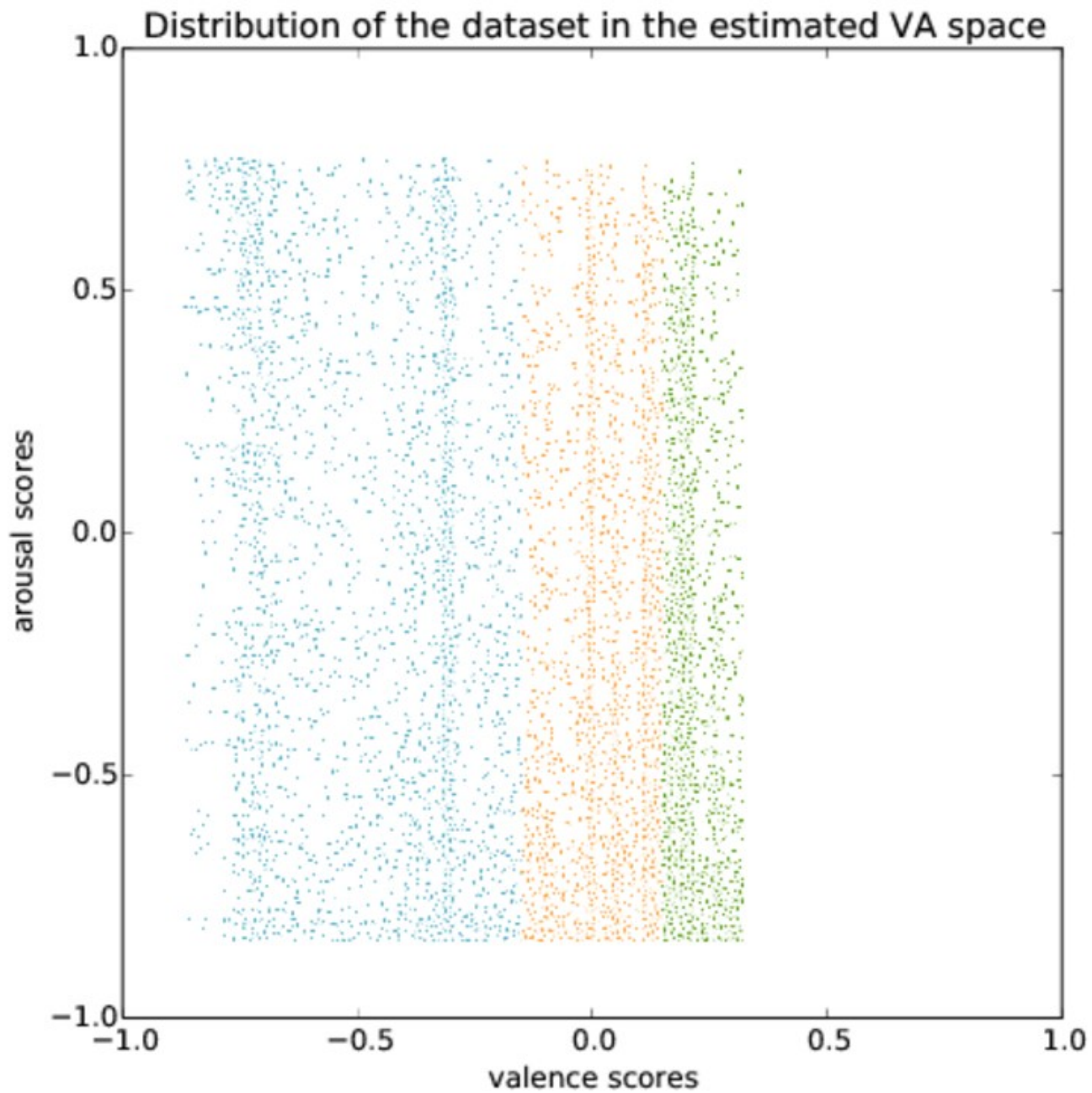
positive

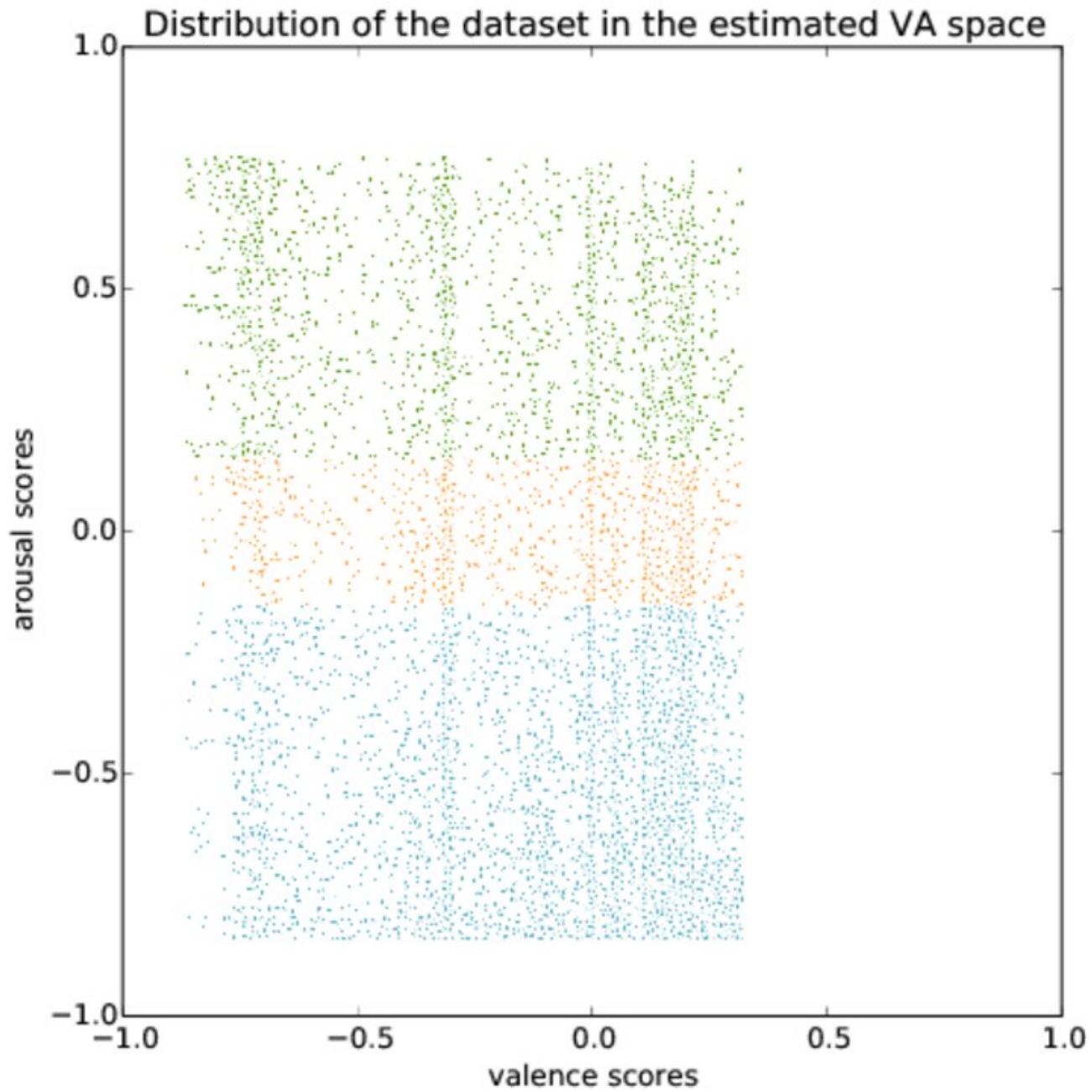
Valence class distribution



Arousal class distribution







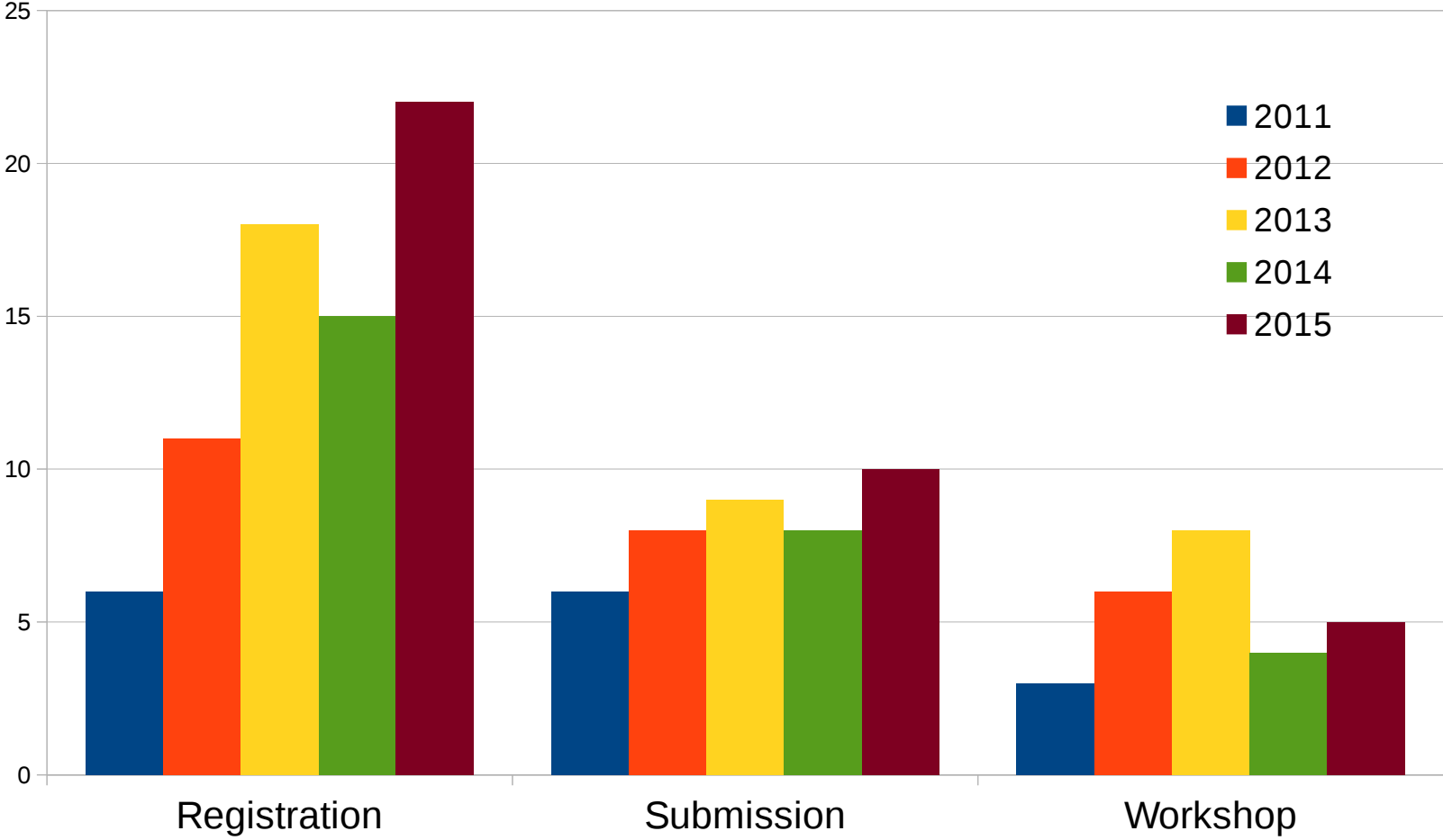
Ground truth annotations: violence

- Violence annotations as in previous years:
 - Two annotation teams
 - Two levels of merging and reviewing
- Videos not originally selected for violence
→ low level of violence
- dev. set: 4.4 % - test set 4.8%
- Realistic situation for general video/movie repository

Evaluation metrics

- Affective impact: **global accuracy** separately for valence and arousal:
 - Proportion of the returned video clips that have been assigned to the correct class (out of three)
- Violence detection: **average precision** using `trec_eval`
 - Average of precisions evaluated at each returned relevant video clip
 - Ordering is critical!

Task participation



Task participation

- Grand total of 87 runs submitted
 - **Affective impact:** 37 runs
 - **Violence:** 50 runs
 - Most teams submitted to both subtasks (8 of 10)
 - 50% new and 50% old teams

Trends this year

Typical system this year:

- Using CNN deep networks for generating features
 - i.e. “external data”
- (Linear) SVM for classification
- Same approach for both subtask, some vary features
- Affect detection as one-against-rest multi-class classification

Trends this year

Features

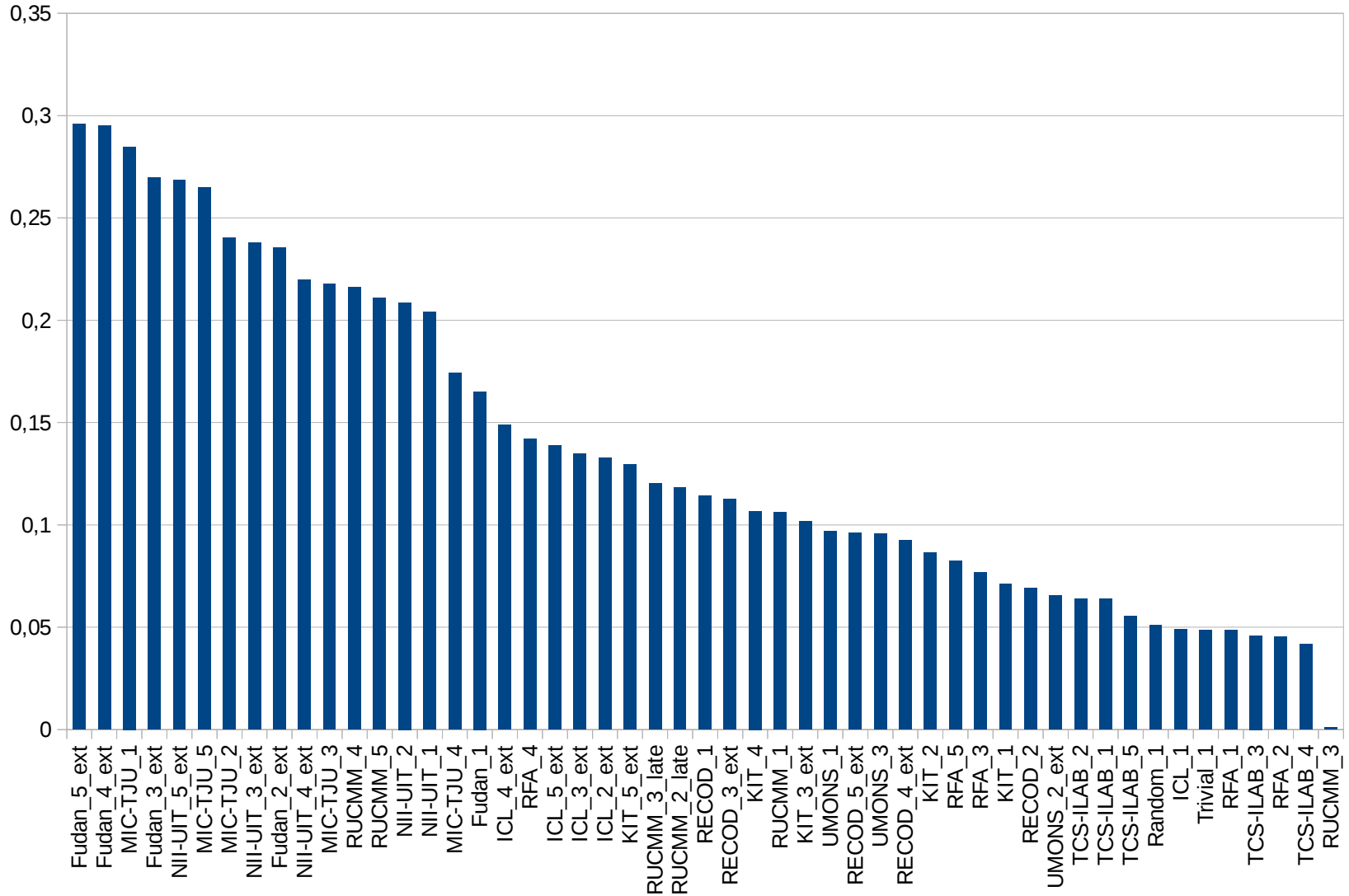
- Typically pre-trained CNNs (ImageNet)
 - Possibly with adaption to MediaEval set
- Two teams: CNNs trained on optical flow maps as temporal feature
- Trajectory-based features common
- GMM Fisher vector encoding popular
- Multi-modality: most use audio, image and motion

Official results – best violence by map

Best run per team	map	R-prec	P10	P100	recall10	recall100
Fudan	0.2959	0.3087	1.0	0.46	0.0435	0.2
MIC-TJU*	0.2848	0.2957	0.8	0.49	0.0348	0.213
<u>NII-UIT*</u>	0.2684	0.313	0.6	0.44	0.0261	0.1913
RUCMM	0.2162	0.2565	0.7	0.38	0.0304	0.1652
<u>ICL</u>	0.1487	0.1913	0.4	0.23	0.0174	0.1
<u>RFA*</u>	0.1419	0.1739	0.3	0.25	0.013	0.1087
KIT	0.1294	0.1783	0.3	0.23	0.013	0.1
<u>RECOD</u>	0.1143	0.1826	0.1	0.18	0.0043	0.0783
<u>UMONS</u>	0.0967	0.1391	0.1	0.13	0.0043	0.0565
TCS-ILAB	0.0638	0.1043	0.0	0.12	0.0	0.0522
<i>Random</i>	0.0511	0.0783	0.1	0.08	0.0043	0.0348
<i>Trivial</i>	0.0486	0.0435	0.0	0.01	0.0	0.0043

* = organising team, underline = present at workshop, R = 230

Results – all violence runs by map



Official results – affective impact: valence

best run	accuracy	negative/neutral/positive
<u>NII-UIT*</u>	42.956	50/43/8
MIC-TJU*	41.947	39/42/18
Fudan	41.779	38/35/27
<u>ICL</u>	41.484	43/37/20
KIT	38.541	36/34/30
<i>Trivial</i>	37.868	0/100/0
<u>UMONS</u>	37.279	47/35/18
<u>RFA*</u>	36.123	17/75/7
TCS-ILAB	35.660	34/37/29
<i>Random</i>	34.651	36/36/28

Official results – affective impact: arousal

best run	accuracy	calm/neutral/active
MIC-TJU*	55.929	92/3/5
<u>NII-UIT*</u>	55.908	99/0/1
<u>ICL</u>	55.719	93/0/7
<i>Trivial</i>	55.551	100/0/0
<u>UMONS</u>	52.439	79/1/20
KIT	51.892	77/8/16
TCS-ILAB	48.949	73/1/27
Fudan	48.844	64/12/23
<u>RFA*</u>	45.038	69/9/22
<i>Random</i>	43.818	62/15/24



Valence results, testset subsets

Testset ACCEDE

best run	acc.	calm/neut/act
NII-UIT*	40.7	49/41/10
MIC-TJU*	40.6	38/42/19
Fudan	40.6	38/35/27
ICL	40.0	40/41/18
KIT	37.8	35/33/32
UMONS	36.4	55/41/3
<i>Trivial</i>	36.1	0/100/0
TCS-ILAB	35.0	32/37/30
RFA*	34.6	18/74/8
<i>Random</i>	34.3	37/36/28

Testset EXTENDED

best run	acc.	neg/neu/pos
NII-UIT*	51.4	52/41/7
ICL	48.3	48/33/20
MIC-TJU*	46.5	43/41/16
Fudan	45.8	41/32/27
<i>Trivial</i>	43.8	0/100/0
KIT	42.1	50/31/19
UMONS	42.0	47/35/18
RFA*	41.3	14/81/5
TCS-ILAB	37.8	37/36/26
<i>Random</i>	36.2	40/35/25

Arousal subsets, testset subsets

Testset ACCEDE

best run	acc.	calm/neut/act
NII-UIT*	64.5	99/0/1
ICL	64.1	100/0/0
<i>Trivial</i>	64.1	100/0/0
MIC-TJU*	63.5	92/3/5
UMONS	59.8	79/1/20
KIT	58.1	76/8/16
TCS-ILAB	55.3	72/0/27
Fudan	53.2	64/13/23
RFA*	50.2	70/9/21
<i>Random</i>	48.2	62/14/23
<i>RandEq</i>	34.1	33/33/33

Testset EXTENDED

best run	acc.	calm/neut/act
<i>Trivial</i>	58.2	0/100/0
NII-UIT*	56.9	19/68/13
KIT	52.4	25/74/2
RFA*	52.4	19/67/14
Fudan	35.4	66/12/22
<i>RandEq</i>	34.5	33/33/33
MIC-TJU*	32.6	91/3/6
UMONS	30.5	22/17/61
ICL	30.4	85/1/14
<i>Random</i>	29.9	61/15/24
TCS-ILAB	29.4	65/13/22

Conclusions

- Violence task
 - Dataset not selected for violence (~4%)
 - Still, teams did quite well!
 - Approaches quite similar to last year
- Affective impact task very hard!
 - Arousal: no improvement over trivial baseline
 - Valence shows slight improvement
 - Change in annotation method very visible
 - Too subjective?

Dataset release

- Testset annotations will be released to participants immediately after workshop
- Planned for October 2015: the dataset extension released publicly on LIRIS-ACCEDE site
 - New video clips
 - Valence/arousal + violence annotations

The future of the Affect Task

- No plans for future violence task
 - Unless someone wants to lead such a task?
- Induced affect task (valence, arousal):
 - LIRIS is planning to continue
- Technicolor is planning to propose a new task:
 - *Predicting image and/or video interestingness*
- Stay tuned for more information later!
 - Join mediaeval-affect mailing list!

Today's program for Affective Impact

- 11:55-12:45: Affective Impact technical retreat in the *Tower room*, open discussion:
 - Lessons learned
 - Future of the task
 - Any other topics ...
- 12:45-14:00: Poster session during lunch