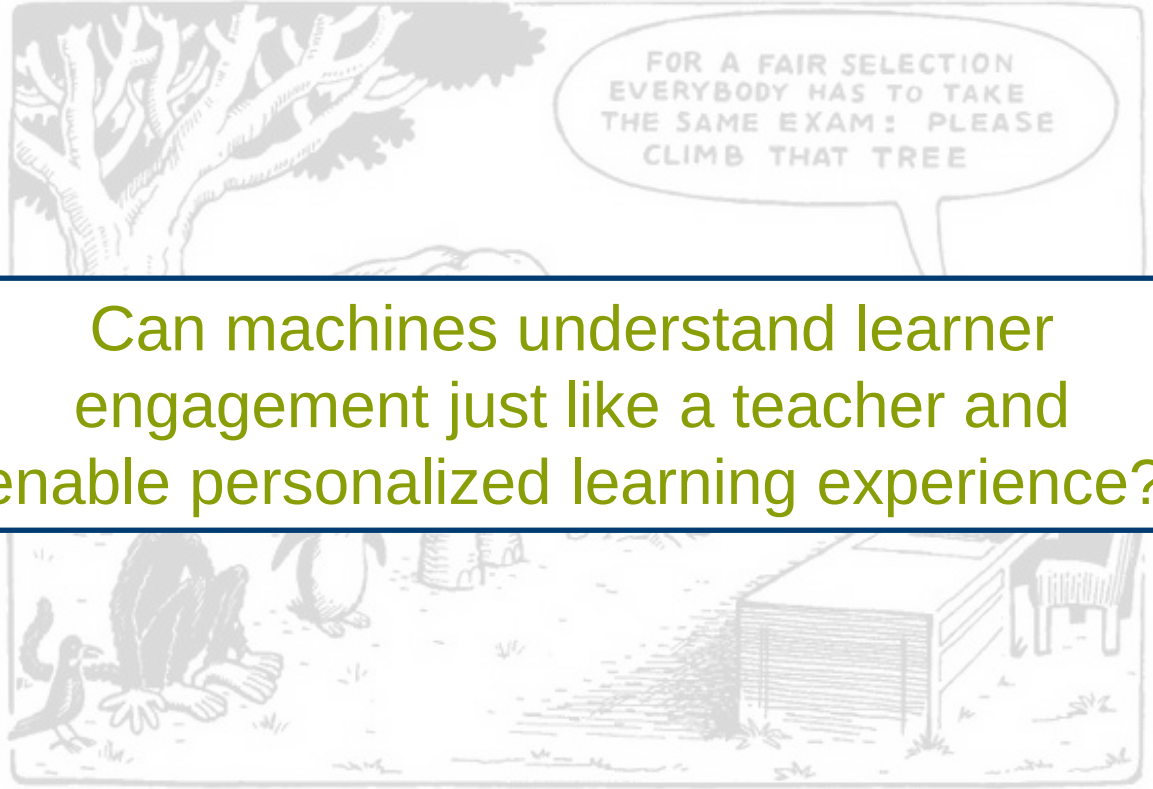
Towards an emotional engagement model:

Can affective states of a learner be automatically detected in a 1:1 Learning Scenario?

Nese Alyuz[1], Eda Okur[1], Ece Oktay[1], Utku Genc[1], Sinem Aslan[1], Sinem Emine Mete[1], David Stanhill[1], Bert Arnrich[2], Asli Arslan Esme[1]

Presenter: **Mustafa Kocaturk[1]**

[1] INTEL CORPORATION
[2] BOGAZICI UNIVERSITY

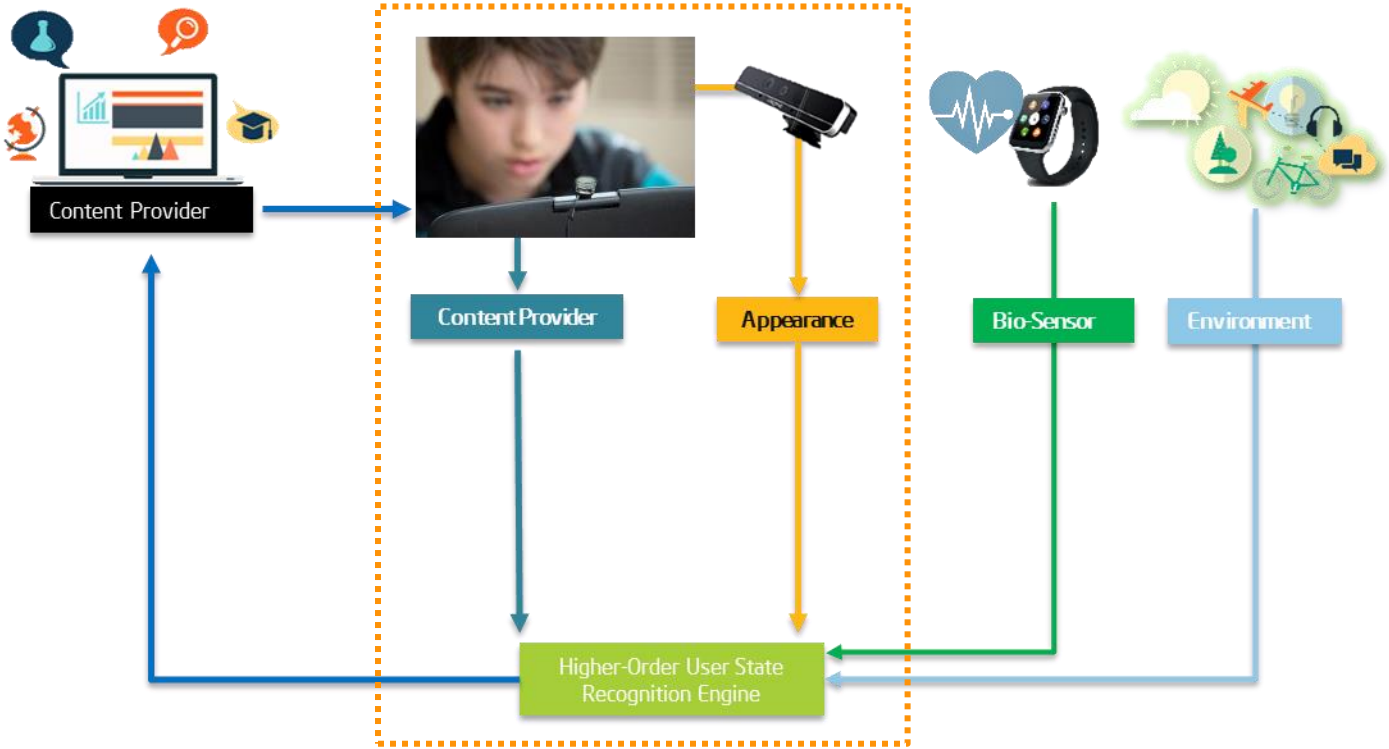


Can machines understand learner engagement just like a teacher and enable personalized learning experience?

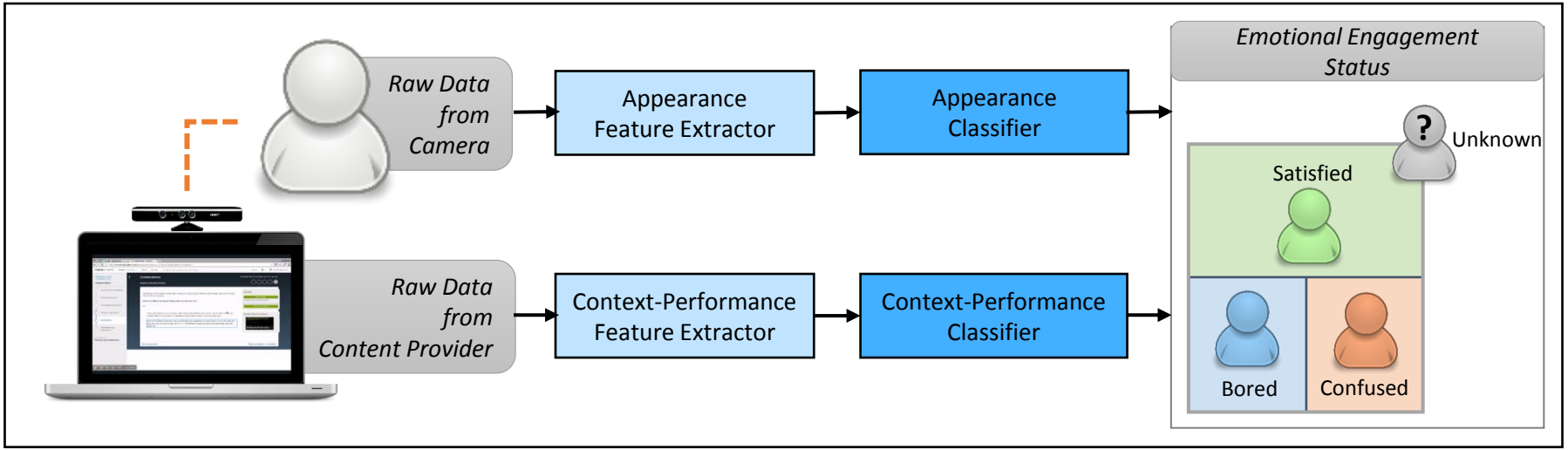
“Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.” Albert Einstein

Picture: http://weknowmemes.com/tag/please-climb-that-tree/

Adaptive Learning System



Generic Emotional Engagement Detection



* S. Aslan, S. E. Mete, E. Okur, E. Oktay, N. Alyuz, U. Genc, D. Stanhill, and A. Arslan Esme, "Human Expert Labeling Process (HELP): Towards a reliable higher-order user state labeling by human experts", in *Int. Conf. on Intelligent Tutoring Systems (ITS) – Workshops*, 2016.

Feature Extraction

Sliding windows of 8-seconds, with 4-seconds overlaps

Modality	Feature Group	Number of Features	Examples
Appearance	Head pose and position	128	median of absolute head center acceleration, standard deviation of head position, etc.
	Facial expressions	32	Number of right-eye raisers per segment, mean of smile, etc.
	Seven basic emotions	28	Mean of anger intensity, number of joyful segments, etc.
Context-Performance	Time related	6	Time from beginning, video/attempt duration, etc.
	Trial related	3	Trial number, number of trials until success, etc.
	Hint related	5	Number of hints used on attempt or question, etc.
	Grade related	7	Grade, correct attempt percentage, etc.
	Other	3	Gender, question number from beginning, etc.

Data Collection

- **Setup:**

  Authentic classroom pilots with 9th grade students

  Optionally offered math course through a public online math learning tool

  Instructional (watching videos) vs. **Assessment** (solving exercises) sections

- **Data:**

  17 one-hour sessions (twice a week), 17 students

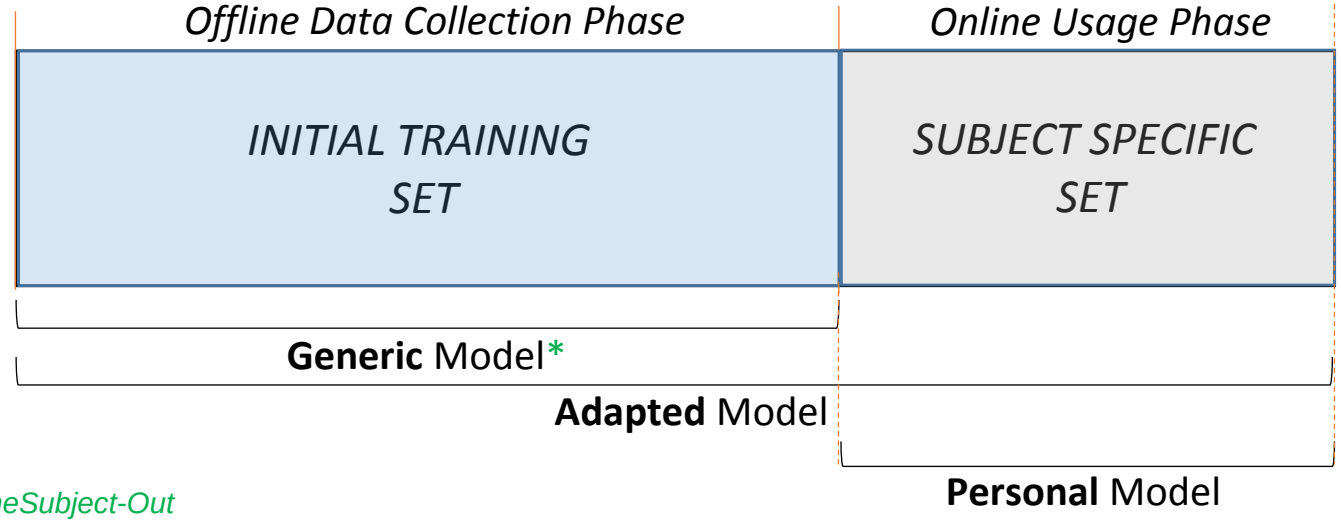
  Human Expert Labeling Process (HELP)* with 8 labelers (5 labelers per instance)

  210 hours of data

* S. Aslan, S. E. Mete, E. Okur, E. Oktay, N. Alyuz, U. Genc, D. Stanhill, and A. Arslan Esme, "Human Expert Labeling Process (HELP): Towards a reliable higher-order user state labeling by human experts", in *Int. Conf. on Intelligent Tutoring Systems (ITS) – Workshops*, 2016.

Experiments

- **Aim:** Need for model personalization
- **Experimental Data:** Using data of nine students (attended sessions twice a week)
- **Experiments:** ***Generic*** vs. ***Adapted*** vs. ***Personal*** Emotional Engagement Model



** Leave-OneSubject-Out*

Classifier Setup

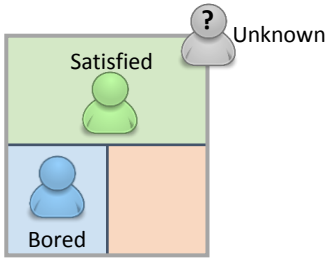
- Training/Test separation: **80%** vs. **20%** of subject specific data
- **Balanced** training sample counts for different classes (10-fold)

- Separate ***Random Forests*** classifiers for:
  - Different modalities: Appearance | Context-Performance
  - Different section types: Instructional | Assessment

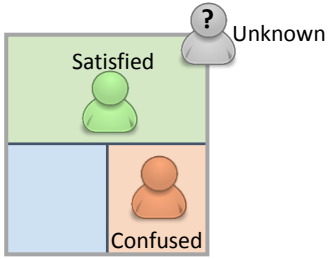
- ***F1 measure*** as the performance criteria:

$$F_1 = 2\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Instructional

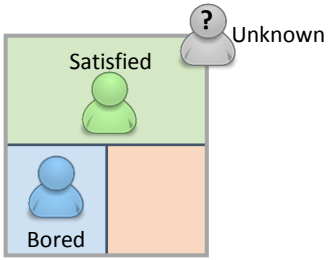


Assessment



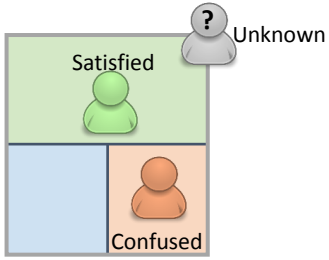
Classification Results

INSTRUCTIONAL



Classes	GENERIC MODEL			ADAPTED MODEL			PERSONAL MODEL		
	Avg. Tr. Size	Appr. (%F1)	C-P (%F1)	Avg. Tr. Size	Appr. (%F1)	C-P (%F1)	Avg. Tr. Size	Appr. (%F1)	C-P (%F1)
Unknown	967	10.73	9.62	1018	24.85	72.97	51	33.04	85.38
Satisfied	967	61.04	55.76	2273	87.63	96.12	1305	89.65	97.18
Bored	967	44.93	39.68	1542	70.91	93.33	575	73.54	94.41
OVERALL	2901	55.79	49.50	4833	85.44	96.13	1931	89.30	97.32

ASSESSMENT



Classes	GENERIC MODEL			ADAPTED MODEL			PERSONAL MODEL		
	Avg. Tr. Size	Appr. (%F1)	C-P (%F1)	Avg. Tr. Size	Appr. (%F1)	C-P (%F1)	Avg. Tr. Size	Appr. (%F1)	C-P (%F1)
Unknown	1886	33.53	27.94	2211	47.21	72.02	324	49.48	72.75
Satisfied	1886	60.58	76.32	2884	83.43	94.04	997	83.79	94.39
Confused	1886	17.12	46.59	2044	37.64	82.05	158	44.04	85.01
OVERALL	5658	48.12	63.41	7139	75.25	90.24	1479	76.37	90.89

Conclusion & Future Work

Findings:

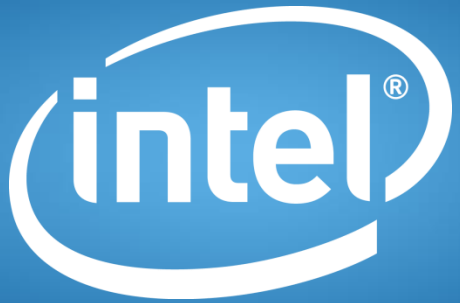
- Appearance is more informative for instructional sections
- Performance-related features makes C-P modality more representative
- Information included in both modalities are person-specific
- Appearance requires more person-specific data

Next?

- Assessment of personalization with self-labels
- New personalization strategies minimizing the need for self-labels
- Fusion of different modalities

Thank YOU

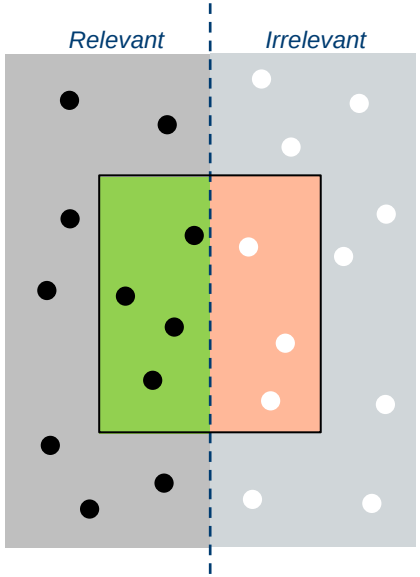
For further questions and comments,
*please contact **Nese Alyuz** at nese.alyuz.civitci@intel.com*

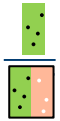

intel

F1 Measure

- ***F1 measure*** as the performance criteria

$$F_1 = 2\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$



$$\text{Precision} =$$ 

$$\text{Recall} =$$