The Hinterland of Emotions: Facing the Open-Microphone Challenge

September 12, 2009

Stefan Steidl[1], Björn Schuller[2], Anton Batliner[1], Dino Seppi[3]

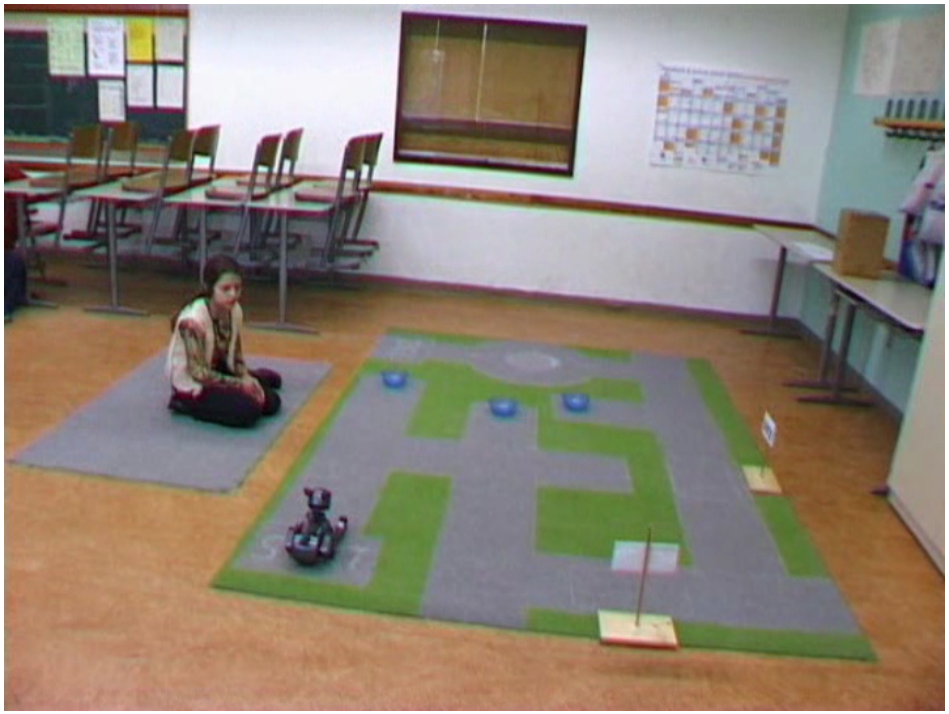
[1]University Erlangen-Nuremberg, Germany

[2]Technische Universität München, Germany

[3]Katholieke Universiteit Leuven, Belgium

Overview

FAU Aibo Emotion Corpus



- children interacting with Sony pet robot Aibo:
  children use their voice to control Aibo
- 51 children at the age of 10-13 year from two schools

FAU Aibo Emotion Corpus (cont.)

- 11 emotion-related states;
  annotated by 5 labellers on the word level
- in this study:
  - 2-class problem: **NEG**ative valence vs. **IDL**e
  - classification of chunks:
    18 216 manually defined syntactic-prosodic chunks
  - heuristic mapping of emotion labels on the word level
    onto single label for the whole chunk
- now available for **scientific, non-commercial use**
  `http://www5.cs.fau.de/FAUAiboEmotionCorpus`

Prototypicality

In our context:

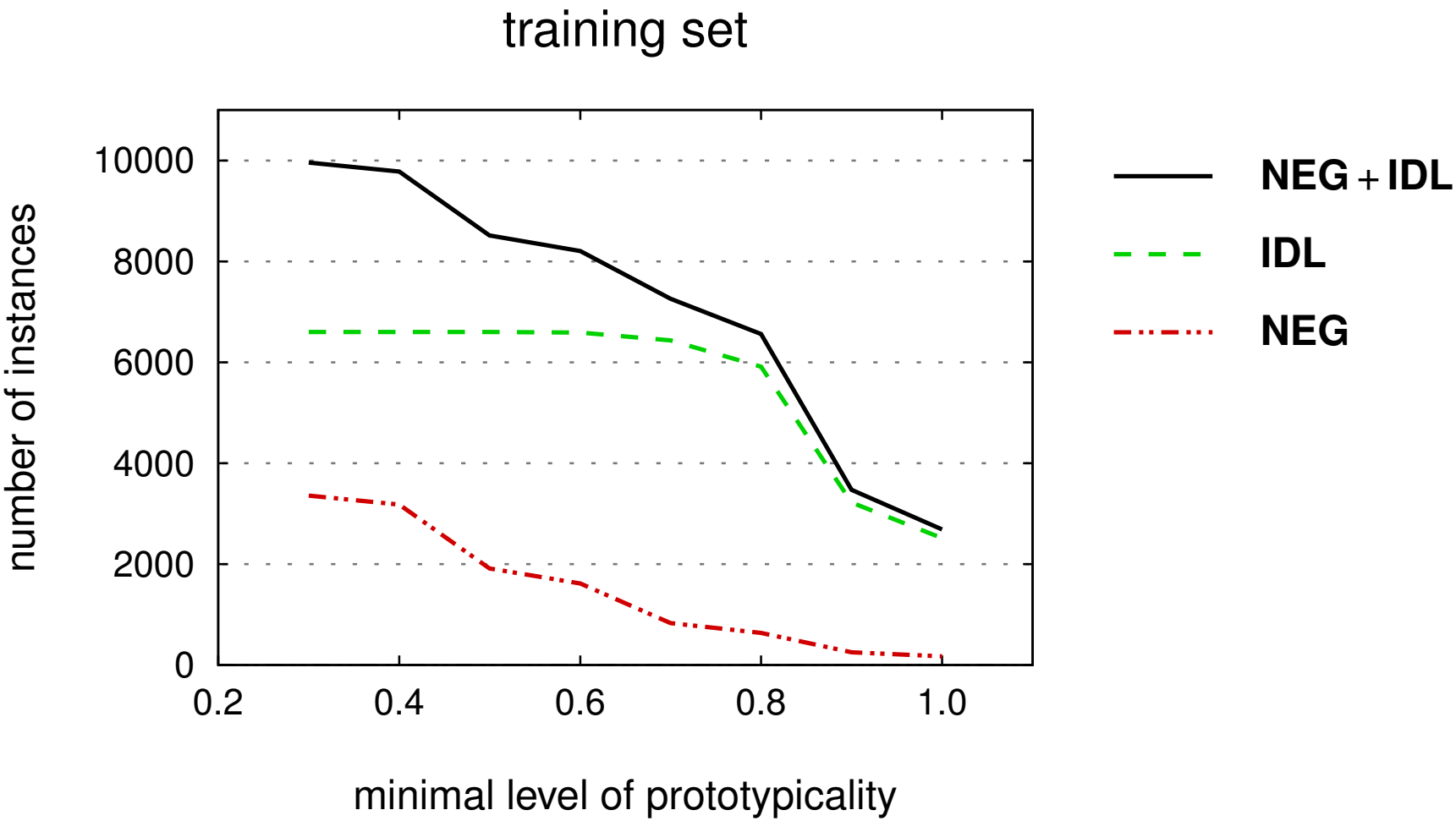
Prototypicality

Proportion of raw emotion labels, i. e. original emotion labels of the five human labellers on the word level,
that match the single emotion label for the whole chunk.

Reasons for low prototypicality

- low inter-rater agreement due to low emotional intensity
- chunks consisting of more than one emotion-related state

Different Levels of Prototypicality



Automatic Emotion Recognition (AER)

Set of acoustic features

- 1 638 features computed with the free software openSMILE
  `http://sourceforge.net/projects/opensmile/`

LLD (26 · 3)	**functionals** (21)
(Δ/ΔΔ) ZCR	mean, abs. mean, centroid
(Δ/ΔΔ) DC, Min, Max	std. deviation, variance
(Δ/ΔΔ) RMS Energy	kurtosis, skewness
(Δ/ΔΔ) LOG Energy	*extremes:*
(Δ/ΔΔ) F0 frequency	value, rel. pos., range
(Δ/ΔΔ) F0 strength	*linear regression:*
(Δ/ΔΔ) F0 quality	offset, slope, MAE, MSE
(Δ/ΔΔ) HNR	*quadratic regression:*
(Δ/ΔΔ) MFCC 0-15	coeff. 1-3, MAE, MSE

Automatic Emotion Recognition (AER) (cont.)

Linguistic features

- based on ASR output
  - vocabulary: 813 words
  - OOV rate: 2.1 % (word fragments)
  - language model: back-off bi-grams
  - 77.48 % word accuracy on test set
- bag-of-n-grams ($1 \leq n \leq 3$)

Automatic Emotion Recognition (AER) (cont.)

Linguistic features

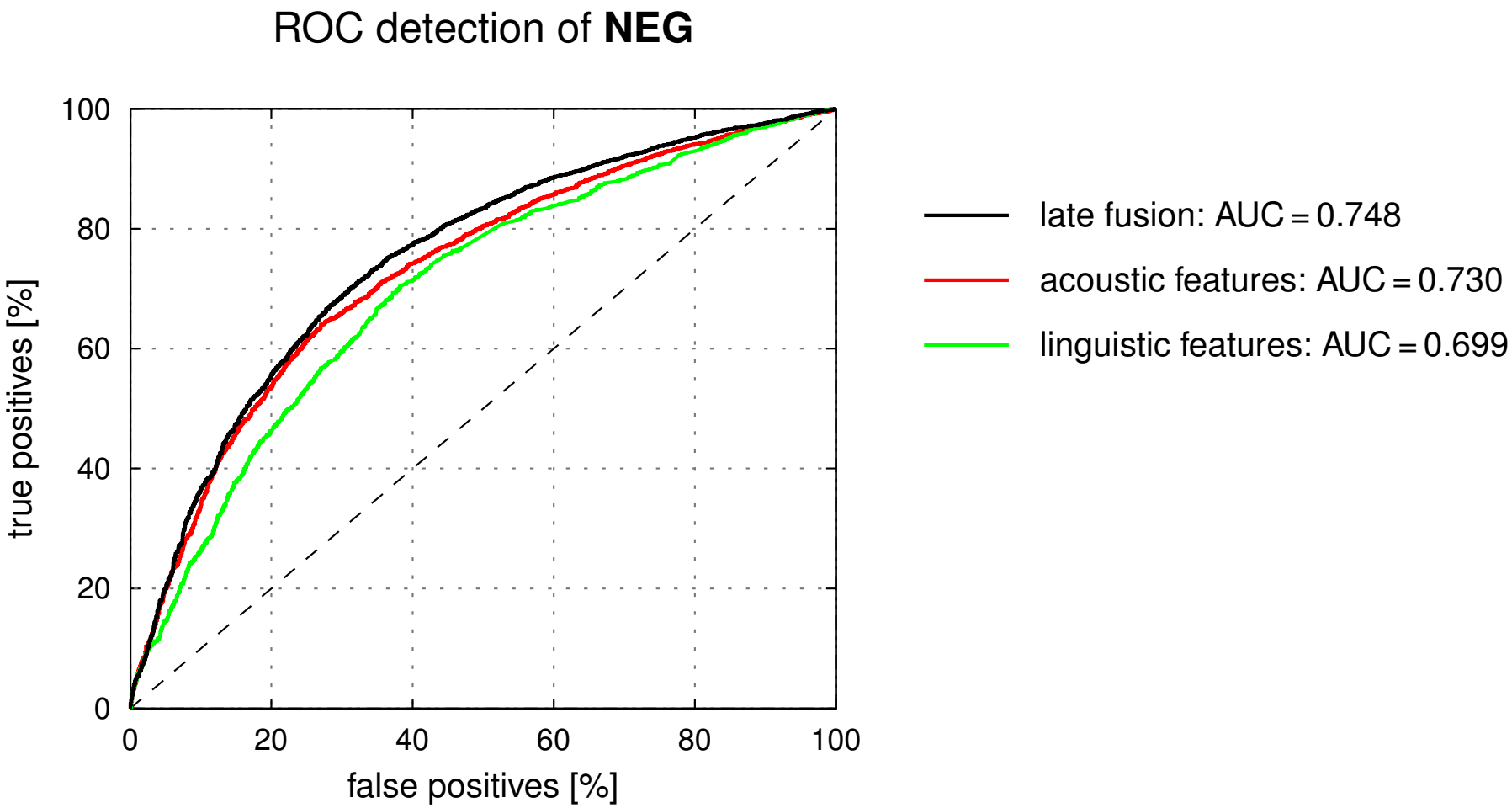
- based on ASR output
  - vocabulary: 813 words
  - OOV rate: 2.1 % (word fragments)
  - language model: back-off bi-grams
  - 77.48 % word accuracy on test set
- bag-of-n-grams ($1 \leq n \leq 3$)

Classification results

average recall [%]	unweighted UA	weighted WA
acoustic features	68.30	65.97
linguistic features	66.05	67.87
late fusion	69.30	71.47

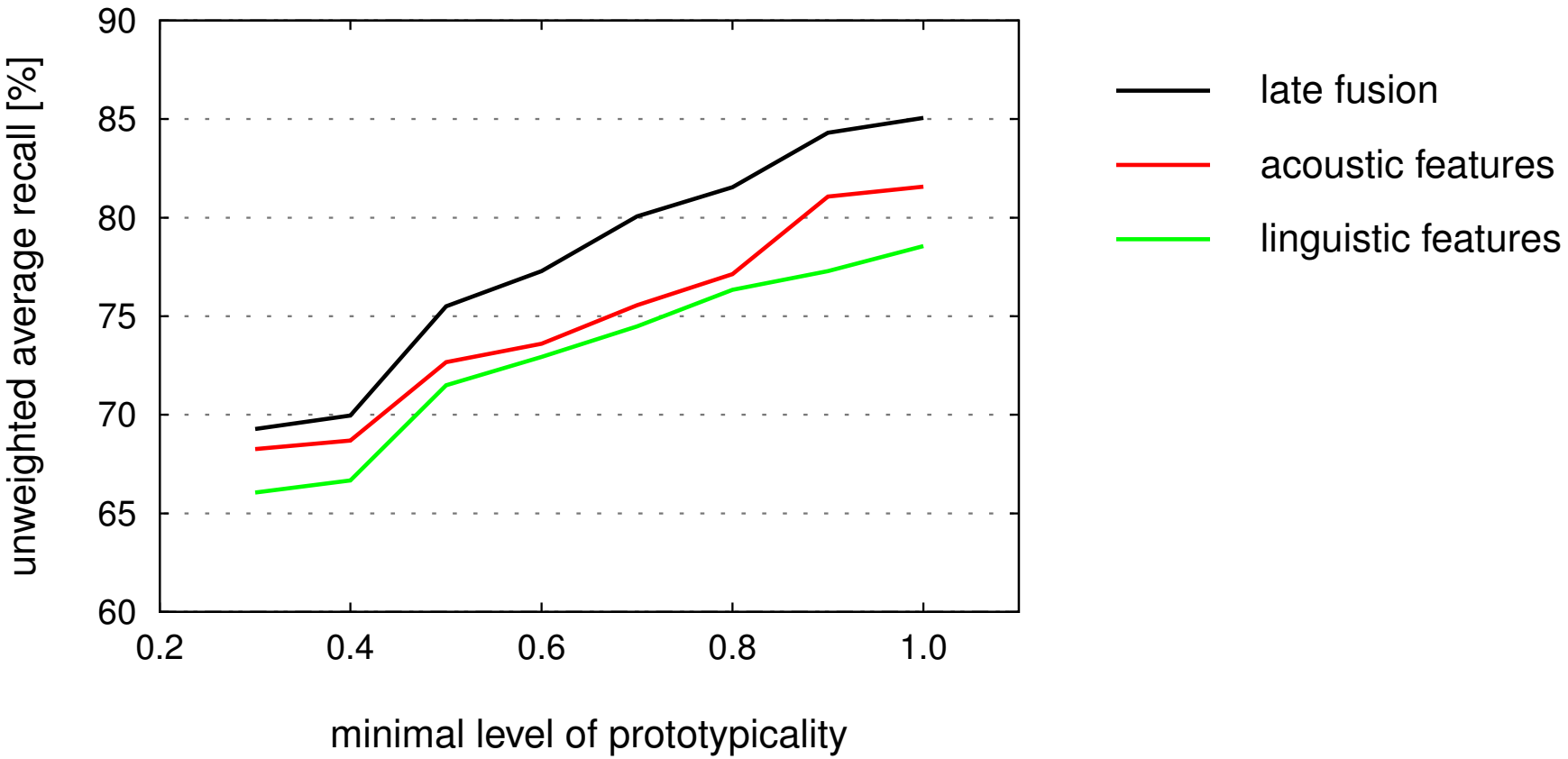
Impact of Prototypicality on AER (cont.)

Receiver operating characteristic (ROC)



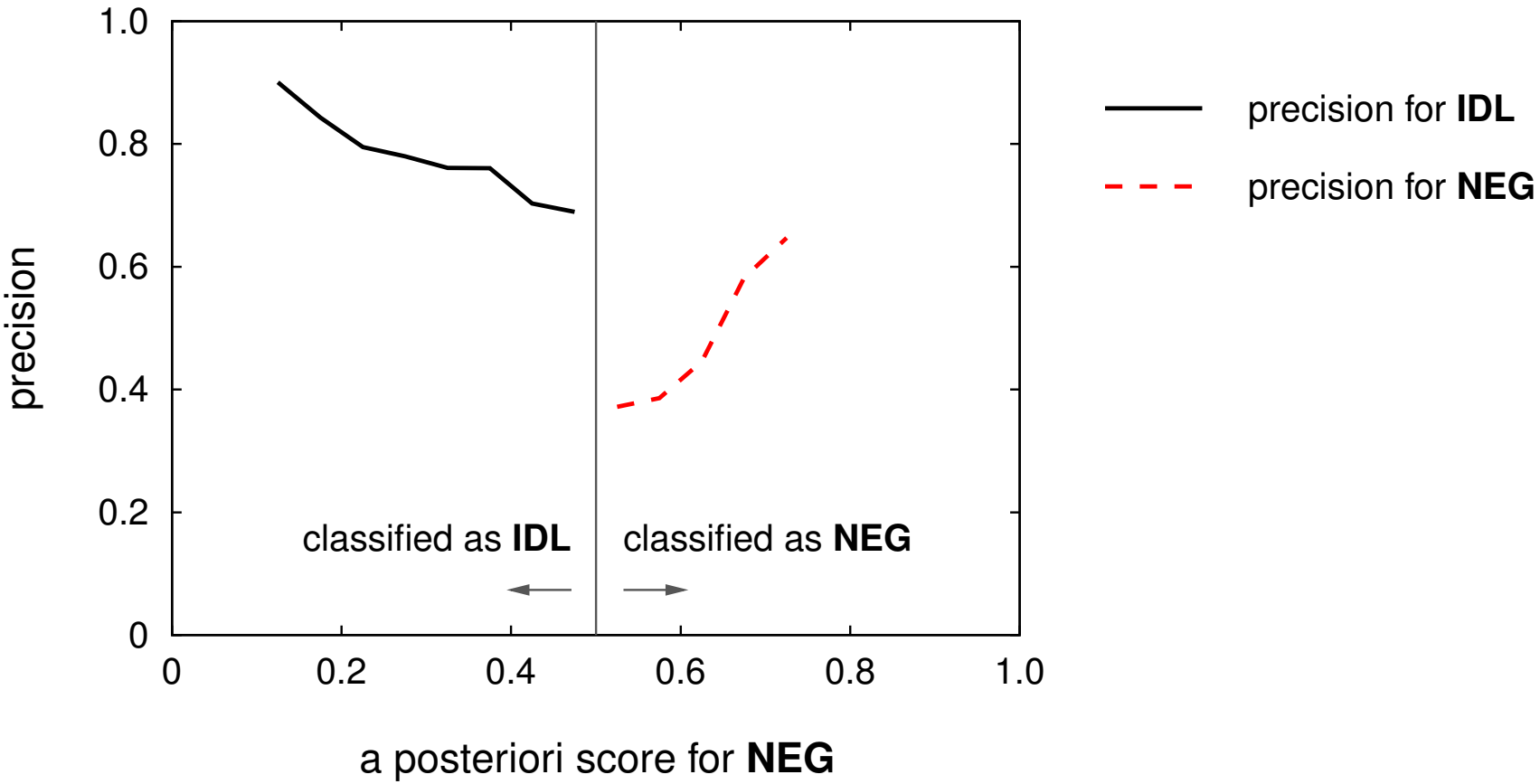
Impact of Prototypicality on AER

Unweighted average recall



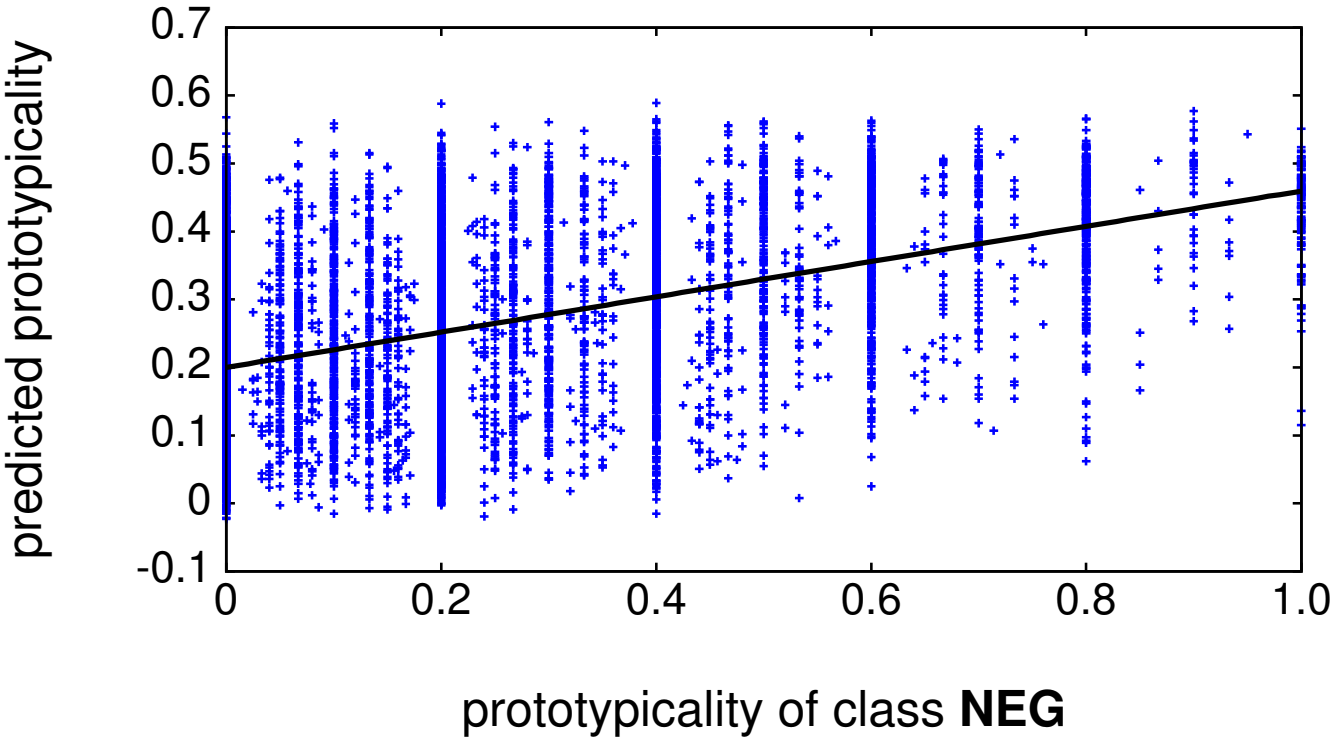
Predicting the Level of Prototypicality

Based on a posteriori scores



Predicting the Level of Prototypicality (cont.)

Based on Support Vector Regression



- correlation coefficient: 0.46
- mean linear error: 0.18

Thank you for your attention!