Spoken Attributes: Mixing Binary and Relative Attributes to Say the Right Thing
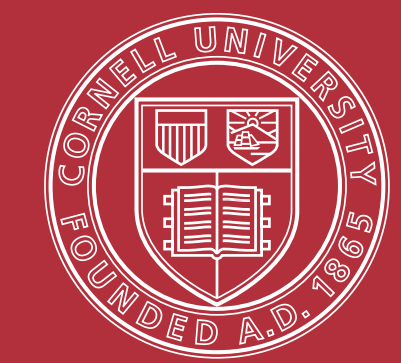
Cornell University

Amir Sadovnik[1]   Andrew Gallagher[1]   Devi Parikh[2]   Tsuhan Chen[1]
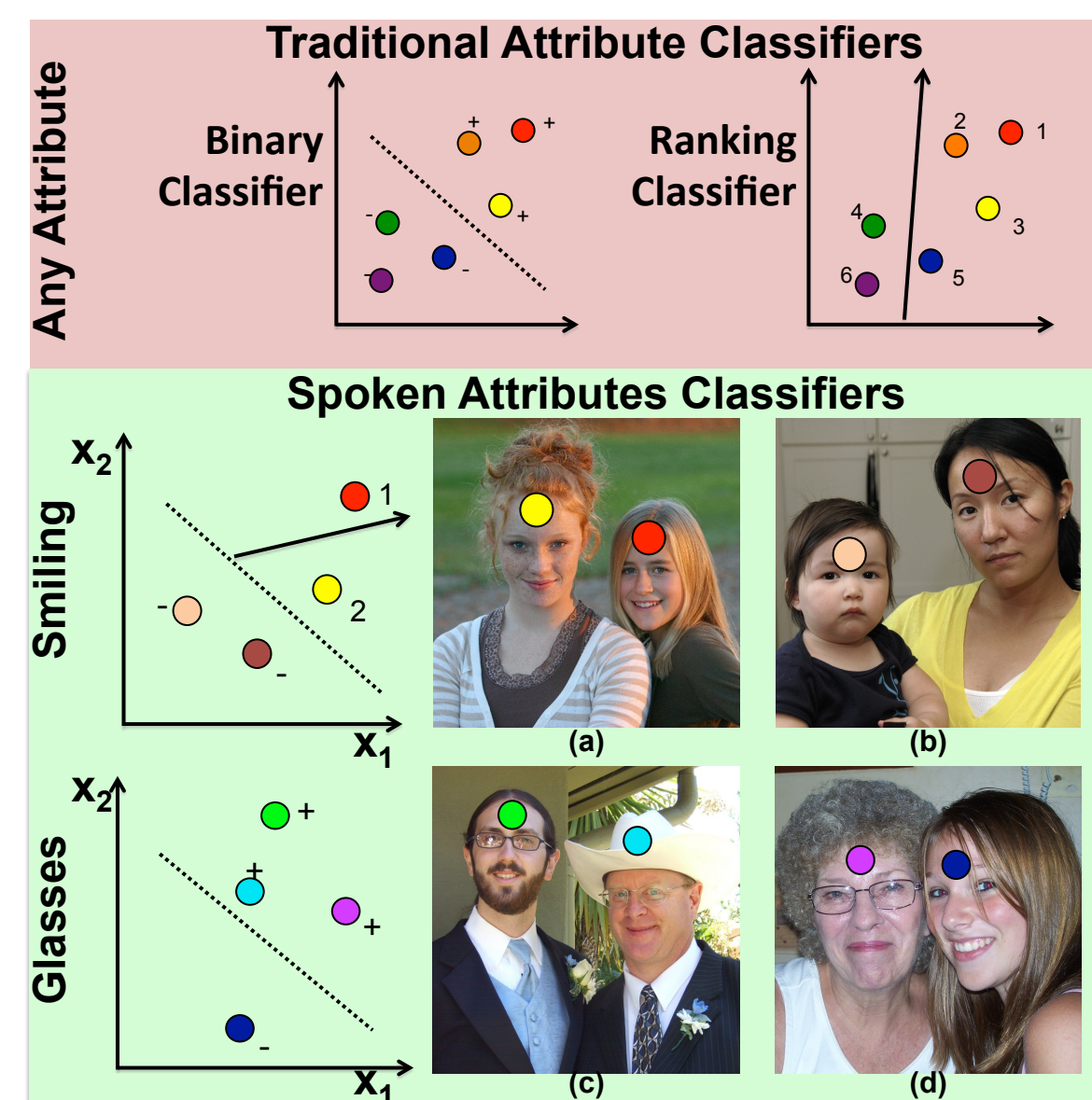
[1]Cornell University   [2]Virginia Tech

VirginiaTech
Invent the Future®

Motivation

- Most previous work has treated attributes as either binary or relative.
- However, this does not model the actual way humans use these attributes in descriptive statements of an image.
- Each attribute might be used as relative or binary depending on situational context
- We model this context for each attribute and learn when it is appropriate to use binary vs. relative

Examples of two different attributes. Smiling is typically used in relative form when both people are smiling, but "has glasses" attribute is never used in relative form.

Understanding Spoken Attributes

For 600 images of pairs of people we collect the following ground truth data:
- **Binary ground truth for each face:**
  - A has/does not have attribute
  - B has/does not have attribute
- **Relative ground truth for each pair:**
  - A has attribute more than B
  - B has attribute more than A
  - A and B have attribute equally
- **Spoken Attribute ground truth for pair:**
  - 🔵 A and B have attribute
  - 🟢 A and B do not have attribute
  - 🔴 A has attribute and B does not
  - 🔴 B has attribute and A does not
  - 🟡 A has attribute more than B
  - 🟡 B has attribute more than A

Spoken Attributes Ground Truth:	Binary Ground Truth			Relative Ground Truth	
	Both	Neither	Only One	Unequal	Equal
Bald	7.65%	76.53%	15.82%	40.18%	59.82%
Beard	9.95%	66.38%	23.67%	36.46%	63.54%
Glasses	12.86%	65.87%	21.27%	21.48%	78.25%
Male	31.79%	17.26%	50.94%	66.73%	33.27%
Smiling	69.69%	7.71%	22.60%	71.27%	28.73%
Teeth Visible	42.71%	19.15%	38.14%	67.65%	32.35%

Both	Neither	Only one	Relative

- Binary expressions cannot express relative relationships when appropriate
- Relative expression "force" users to select a relative statement inappropriately
- Appropriate usage of relative and binary statement are attribute-dependent
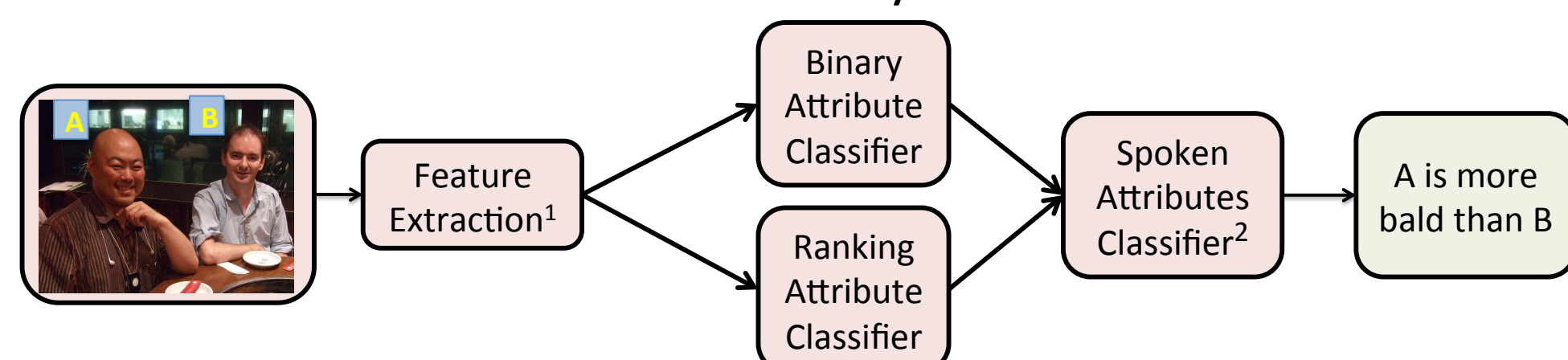
Prediction Results

We use leave-one-out cross validation and compare with the following baselines:
- **Raw Feature:** skip first layer classifiers, and concatenate raw feature vectors of pair.
- **Difference:** skip first layer classifiers, and take difference of raw feature vectors of pair.
- **Relative:** can only output relative statement (or majority binary statement if below optimal threshold).
- **Binary:** Can only output a binary statement

	Bald	Beard	Glasses	Male	Smiling	Teeth Visible	Average Accuracy
Spoken Attributes	0.8231	0.7976	0.8958	0.7470	0.5421	0.5744	0.7300
Raw Feature	0.7906	0.7273	0.8118	0.7051	0.5220	0.5884	0.6909
Difference	0.7834	0.7118	0.7160	0.5567	0.4341	0.3958	0.5996
Relative	0.7762	0.6775	0.6521	0.2723	0.4469	0.2278	0.5088
Binary	0.8321	0.7513	0.8571	0.7500	0.4542	0.5447	0.6982
Majority Guess	0.7762	0.6724	0.6521	0.3316	0.3608	0.2277	0.5035

0%   50%   100%

Our Approach

- We build a classifier, that selects the **most accurate statement** to describe a pair of faces.
- We model this as a multiclass classifier whose output is one of 6 possible statements (4 binary, 2 relative).
- We use a 2-layer classifier, where the first layer consists of a binary SVM and a rank SVM trained on binary and relative data.

[1]21504 feature vector composed of densely sampled SIFT descriptors at different grid cells and scales. B. G. H. Chen, A. Gallagher. What's in a name: First names as facial attributes. In Proc. CVPR, 2013.

[2]Input to Spoken Attributes Classifier:
(binary score of A, binary score of B, relative score of A, relative score of B, difference of relative score)
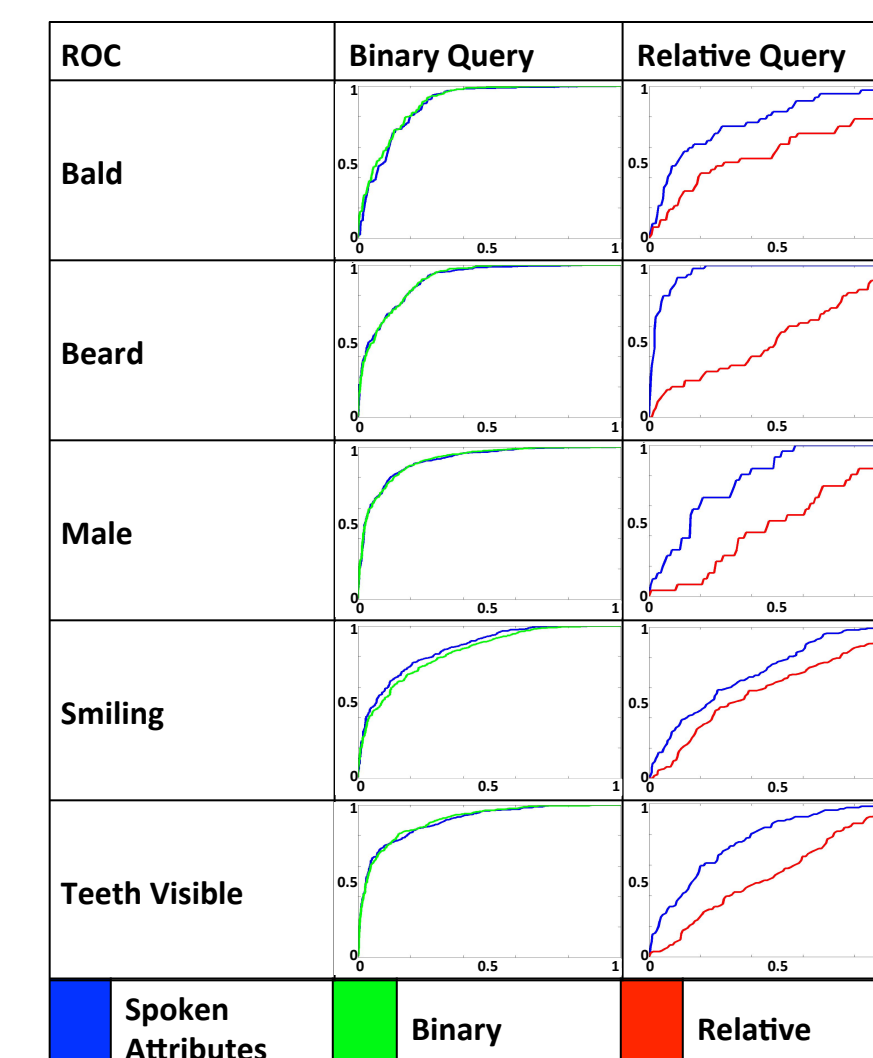
Reading Between the Lines

Given a spoken attribute, can we infer the correct binary statement?
Some attributes are only used in relative form when both people have the attribute.

	Beard	Teeth
Binary Classifiers	0.43	0.84
Binary Trained for SA	0.81	0.92
Most Common For SA	0.82	0.90

	Beard		Teeth Visible	
	A	B	A	B
Binary Prediction	No Beard	Beard	Teeth not visible	Teeth visible
Spoken Attribute Statement	Person B is more bearded than person A		Person B's teeth are more visible than person A's	
Binary Prediction With Statement	Beard	Beard	Teeth visible	Teeth visible

Searching for the Right Image

Given a description, can we return better search results using our spoken attributes classifier score?

Spoken Attributes	Binary	Relative

Examples

Spoken attribute is a relative statement				
Binary	Both People's teeth are visible	Both People are bald	Person A is bearded and person B is not	Both People are smiling
Relative	Both People's teeth are visible	Neither person is bald	Neither person is bearded	Both People are smiling
Spoken	Person B's teeth are more visible than person's A's	Person A is more bald than person B	Person A is more bearded than person B	Person A is more smiling than person B

Spoken attribute is a binary statement				
Binary	Person B has glasses and person A does not	Neither person is smiling	Person A is male and person B is not	Neither person has glasses
Relative	Person B has glasses more than person A	Person A is smiling more than person B	Person A is more masculine than person B	Person B has glasses more than person A
Spoken	Person B has glasses and person A does not	Neither person is smiling	Person A is male and person B is not	Person B has glasses and person A does not

Spoken attribute is wrong				
Binary	Person B has a beard and person A does not	Both people are smiling	Neither person is bald	Both people's teeth are visible
Relative	Neither person is bearded	Both people are smiling	Neither person is bald	Person B's teeth are more visible than person A's
Spoken	Person B is more bearded than person A	Neither person is smiling	Person A is bald and person B is not	Both people's teeth are visible
Ground Truth	Person B has a beard and person A does not	Person A is smiling more than person B	Neither person is bald	Person B's teeth are more visible than person A's