Guess who? - Classification of facial attributes

Anna Lindahl
Project report for Embodied and Situated Language Processing course
University of Gothenburg
annanlindahl@gmail.com

Mehdi Ghanimifard
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg
mehdi.ghanimifard@gu.se

Abstract

(AL) Inspired by the game "Guess who", in this project we use neural embeddings of faces to ground their visual features in facial attributes. The faces are taken from the LFW dataset(Huang et al., 2007) with attributes generated by classifiers from Kumar et al. (2009). The embeddings are generated using the OpenFace package (Amos et al., 2016). The embeddings are then used to train a classifier for each attribute. Although the accuracy varies on the classifiers, we argue that embeddings are suitable for these kinds of tasks. To implement the classifiers we design a simple webbased game with our own parser.

1 Introduction & Motivation (AL)

In the game "Guess who?", two players take turns describing the face of a person to the other. They are both presented with the same selection of faces, and the goal is to guess which person the other player describes, based on the description of the visual features of the person.

Inspired by this game, this project aims to classify visual features in faces and connect them to natural language and attributes. To implement the classification we create an adapted version of the Guess who- game, where the user is shown ten images of faces. He or she then has to choose one face and describe it to the computer, which has to guess which face the user is thinking of.

To ground the visual features in attributes, we use neural embeddings made from a pre-trained model. The embeddings reduce the dimensionality which simplifies the classification task.The model is originally made for facial recognition, so the embeddings can be compared to each other, with the distance between them indicating the difference between two faces. This property makes the embeddings suited for our classification task.

2 Related works (MG)

Learning the connection between describable facial attributes and perceptual features in the image is an investigated subject in computer vision (Kumar et al., 2009, 2011). In our project, we partially followed that direction. Instead of using older methods for feature extraction, we took high level representations of the faces based on the face recognition project (Amos et al., 2016), and we trained classifiers for each attributes.

In a highly related project Matuszek et al. (2012) explored a possible model of grounded attribute learning. This learning model depends on a probabilistic semantic parser which recognizes the linguistic categories and if they are one of the given attributes (color, shape and spatial relations), then it learns the grounding relation of these attribute and perceptual representations based on a set of dataset of descriptions of scenes that they collected. In our project, we learn grounded attributes using a dataset collected from human judgments (Kumar et al., 2009), then during the game our parser recognizes these attributes in natural language. Our classifier figures out in which picture they can be grounded.

As our parser depends on a common sense information including the synonyms and antonyms of each attribute, we only have a manually extracted knowledge base. It worth to mention, that there are projects aimed to extract these information either from textual data (Mitchell et al., 2015) or from relation between language and vision (Zhu et al., 2014; Vedantam et al., 2015). Perhaps, this can be one of the directions for future study in lined with this project.

3 Implementation (MG)

The implementation is done in three phases. First, we train all classifiers for each facial attributes. Then, we designed a simple web interface mock-

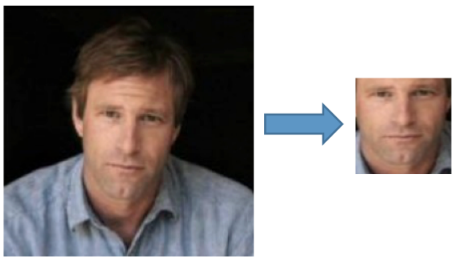

Figure 1: Example of cropping and aligning an image.

up and considered different interactive scenarios for the game, with Wizard of OZ methodology (Kelley, 1984). Finally, based on the time limits for the project, we chose one of the dialog scenarios and built a simple parser for the game. The details of each part are explained in details as following:

3.1 Classification of facial features (AL)

For our dataset we used images of faces from the database "Labeled faces in the wild" (Huang et al., 2007). The database was originally put together for the study of facial recognition, and consists of 13233 images of faces of 5749 persons. These images was also used for the game.

To extract embeddings of the images, a pre-trained deep neural net model from the OpenFace package (Amos et al., 2016) was used. First, the faces were identified and the images were aligned based on the location of the nose and the outer edge of the eyes, and then cropped so that only the face were left, see Figure 1. This was done with dlib and OpenCV, via OpenFace. The cropping of the images was necessary for being able to extract useful embeddings. Around 60 images was skipped due to no face detected, mostly because the person was wearing a cap that was covering parts of the face.

After aligning the images the embeddings were extracted with the OpenFace model "nn4.small2.v1". The embeddings have 128 dimensions and has the property that the distance between them represent the likeness of the faces. This enables us to do the classification simple and with short training time.

For facial attributes we used classified attributes from classifiers from Kumar et al. (2009). The classification is already done, and it's based on images described with attributes by humans. There are 73 facial attributes, such as "Male","Blond hair" or "Big nose".[1] There are also attributes related to clothing or image quality, such as "wearing glasses" or "color photo". For every image there is a vector representing each attribute with a positive or negative value, indicating the presence or absence of that attribute. To be able to compare the images with each other we changed the the values to 1 and -1 respectively. A few images didn't have a corresponding attribute vector, those were removed from the dataset.

The embeddings and their respective attribute values was used to train a classifier for each attribute. We used scikitlearn's Linear SVC, a support vector classifier. The result was 73 classifiers for classifying the absence or presence of an attribute. We used a 80:20 distribution for training and test data.

3.2 Parser (AL)

To be able to parse a description of a person to attributes in the game, a simple parser was constructed. The string input from the user is processed in a few steps to identify bigrams that will signify an attribute, for example "blond hair". These bigrams are put together with a "_" and then the sentence is tokenized. The tokens are then compared to our knowledge base,here consisting of two dictionaries of synonyms and antonyms for each attribute. An example of an entry is blond hair:['blond hair', 'blond_hair', 'blondish_hair','blond','blonde']. For each input, the parser returns a vector with 73 entries, 1 for the attributes it detected in the description, -1 for the negation of the attributes it detected (e.g. "not blond") and 0 for the attributes that weren't in the description.

3.3 The Guess Who -game (AL)

The game is implemented in a webserver-style, at this point not all possible features of the game are finished, due to time restrictions. See figure 2 for an example of the game interface.

3.3.1 Game Design (AL)

The game starts with the user providing a username, he or she is then presented by 10 randomly generated images from the testset. The faces are shown with a green background. The user is asked

[1] see appendix A for a list of the attributes.

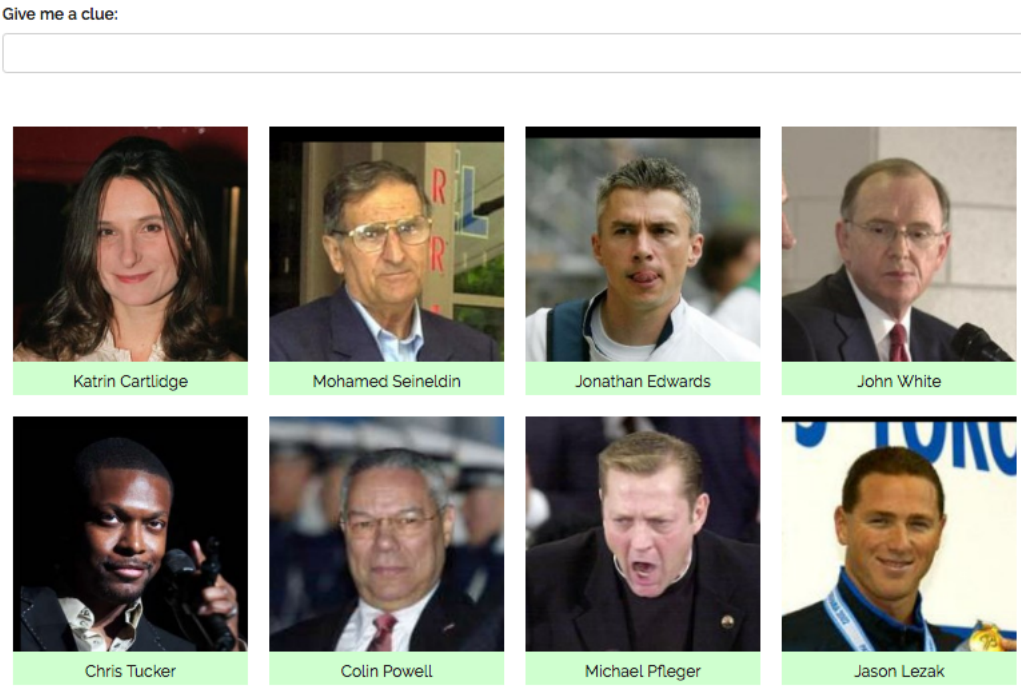

Figure 2: The game interface

to provide description a face. The system then incrementally parses every incoming description and updates the context, the ten faces, based on the clue.

The update works as follows: the parser returns an attribute vector based on the clue, as described previously. This vector is combined with the vector with the previous clues, which will only be a vector of 73 zeros if no relevant information has been provide before. No increase of the values of the vectors occur,the possible values are -1,0 and 1. If there is contradicting information about an attribute in the previous clues and the current clue, the value of that attribute will be reverted to 0. The images are classified and the result from the classification is compared to the clues. If a face has an attribute which is not consistent with the clue, it is removed from the list of possible guesses and changed to white background.

There is currently no end in the game, although it will be obvious for the user if the game has only one person left to guess, this will be the only face with a green background.

3.3.2 Web interface (MG)

The web interface first time was designed as a prototype to present the idea of how the game should look like. Then, after deciding over the scenario of the game, it got fully implemented using two different parts:

1. The webserver using Flask micro-framework in Python (Grinberg, 2014). This part wraps the core logic, including the classifier and the parser with the server interface, which can communicate with the graphical user interface.

2. The javascript dialog manager works within the graphical interface, which handles simple user interactions and pass it on to the webserver for processing[2].

4 Results

The evaluation of the attribute classifiers are done quantitatively. The accuracy, precision, recall and F-score of this can be seen in Appendix B. Among the best classified attributes are ”Male”, ”Eyes open” and ”Round jaw”. The attributes with lowest scores were ”Baby”,”Blond Hair” and ”Heavy Makeup”. Many of the attributes had a high accuracy but low F-score due to low recall or precision.

The success of the game will of course depend on the accuracy of the parser and the classification. Currently, we haven’t presented the game to any users.

[2]The source code of the demo is available at `https://github.com/nekonell/guesswho_lfw`

5 Discussion (AL)

The cropping of the pictures were necessary to be able to classify and compare the pictures, but obviously this is not ideal since this will affect the accuracy of the attribute classifiers. The low scores for attributes such as "Blond Hair" is expected, since that attribute simply isn't present in the images. Though, one could expect that some clues to attributes would be present in faces, such as dark eyebrows might indicate dark hair. We didn't find this connection in our data for this attribute.

A surprising find is that the F-score for "wearing necktie" is very high. This is could be due to some bias in the dataset, most of persons might be wearing necktie. The bias of the dataset is important to take into account for other features, the dataset is primarily made up of people labeled "white". In the attributes there are also two features which are highly subjective, "Attractive man" and "Attractive woman".

The attributes which had the highest scores were attributes present in the faces, which indicates that the embeddings are suited for this kind of task, although preferably with more suited attributes or less cropped pictures.

It's important to remember that even the correct attribute could differ from the actual case, since they are coming from classification based on human judgements, not direct human judgements. Ideally, we would have attributes based directly on human judgements, but we don't have the resources for that.

Another problem is the values of the attribute vectors. As mentioned earlier, they originally were in a range of numbers, with negative and positive numbers indicating absence and presence of an attribute. The higher/lower the number, the stronger the indication. How these numbers were generated and the relation between them was unclear, so we had to change them to binary values. This removes the gradation in the classification.

6 Further work (AL)

There are many directions in which one could extend this project. The classification could be improved, as mentioned earlier, with new attributes and non-cropped embeddings. These embeddings would have to come from another model though, since OpenFace don't work well without the cropping. There are a few other models designed for facial to recognition to try. If one had the resources to collect data and judgements designed for this task, it would be ideal.

For the game, there are also many improvements. At present, the attributes are written by hand. As mentioned previously, more sophisticated attempts exist to construct a knowledge base. One way could be to extend the knowledge base incrementally based on the input of the users. One could also implement attributes not strictly related to visual features, such as the occupation of the person. (The faces from LFW are of celebrities.) The update function could also be more advanced, for example taking into account the probability of co-occurrence of the attributes.

References

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). *OpenFace: A general-purpose face recognition library with mobile applications*. Technical report, CMU-CS-16-118, CMU School of Computer Science.

Grinberg, M. (2014). *Flask Web Development: Developing Web Applications with Python*. " O'Reilly Media, Inc.".

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*. Technical Report 07-49, University of Massachusetts, Amherst.

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems (TOIS)*, 2(1), 26–41.

Kumar, N., Berg, A., Belhumeur, P. N., & Nayar, S. (2011). Describable visual attributes for face verification and image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10), 1962–1977.

Kumar, N., Berg, A. C., Belhumeur, P. N., & Nayar, S. K. (2009). Attribute and simile classifiers for face verification. In *2009 IEEE 12th International Conference on Computer Vision* (pp. 365–372).: IEEE.

Matuszek, C., FitzGerald, N., Zettlemoyer, L., Bo, L., & Fox, D. (2012). A joint model of language and perception for grounded attribute learning. *arXiv preprint arXiv:1206.6423*.

Mitchell, T., Cohen, W., Hruschka, E., Talukdar, P., Betteridge, J., Carlson, A., Dalvi, B., Gardner, M., Kisiel, B., Krishnamurthy, J., Lao, N.,

Mazaitis, K., Mohamed, T., Nakashole, N., Platanios, E., Ritter, A., Samadi, M., Settles, B., Wang, R., Wijaya, D., Gupta, A., Chen, X., Saparov, A., Greaves, M., & Welling, J. (2015). Never-ending learning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15)*.

Vedantam, R., Lin, X., Batra, T., Lawrence Zitnick, C., & Parikh, D. (2015). Learning common sense through visual abstraction. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2542–2550).

Zhu, Y., Fathi, A., & Fei-Fei, L. (2014). Reasoning about object affordances in a knowledge base representation. In *European Conference on Computer Vision* (pp. 408–424).: Springer.

A List of attributes

This is the list of attributes collected by Kumar et al. (2009). We based our parser and our classifiers on these attributes.

Male
Asian
White
Black
Baby
Child
Youth
Middle Aged
Senior
Black Hair
Blond Hair
Brown Hair
Bald
No Eyewear
Eyeglasses
Sunglasses
Mustache
Smiling
Frowning
Chubby
Blurry
Harsh Lighting
Flash
Soft Lighting
Outdoor
Curly Hair
Wavy Hair
Straight Hair
Receding Hairline
Bangs
Sideburns
Fully Visible Forehead
Partially Visible Forehead
Obstructed Forehead
Bushy Eyebrows
Arched Eyebrows
Narrow Eyes
Eyes Open
Big Nose
Pointy Nose
Big Lips
Mouth Closed
Mouth Slightly Open
Mouth Wide Open
Teeth Not Visible
No Beard
Goatee

Round Jaw
Double Chin
Wearing Hat
Oval Face
Square Face
Round Face
Color Photo
Posed Photo
Attractive Man
Attractive Woman
Indian
Gray Hair
Bags Under Eyes
Heavy Makeup
Rosy Cheeks
Shiny Skin
Pale Skin
5 o' Clock Shadow
Strong Nose-Mouth Lines
Wearing Lipstick
Flushed Face
High Cheekbones
Brown Eyes
Wearing Earrings
Wearing Necktie
Wearing Necklace

B Classification Results

Attribute	Accuracy	F-score	Precision	Recall
Male	92.9	95.39	93.59	97.27
Asian	95.8	67.46	79.17	58.76
White	89.08	93.09	89.58	96.88
Black	97.86	69.57	68.09	71.11
Baby	99.77	0.0	0.0	0.0
Child	95.88	0.0	0.0	0.0
Youth	86.41	51.37	67.63	41.41
Middle Aged	89.69	5.59	44.44	2.99
Senior	85.11	59.96	69.19	52.9
Black Hair	88.93	33.18	65.45	22.22
Blond Hair	95.95	43.01	71.43	30.77
Brown Hair	65.95	40.53	53.9	32.48
Bald	89.47	0.0	0.0	0.0
No Eyewear	84.58	90.65	86.18	95.61
Eyeglasses	87.48	55.43	74.45	44.16
Sunglasses	98.63	0.0	0.0	0.0
Mustache	93.66	70.25	86.73	59.04
Smiling	74.12	67.81	76.77	60.71
Frowning	72.6	76.82	71.09	83.57
Chubby	73.44	56.06	62.89	50.57
Blurry	84.05	0.0	0.0	0.0

Harsh Lighting	72.75	36.14	70.14	24.34
Flash	82.37	6.48	57.14	3.43
Soft Lighting	69.92	82.14	70.56	98.26
Outdoor	67.71	57.32	69.61	48.71
Curly Hair	68.78	75.98	69.57	83.7
Wavy Hair	69.92	63.04	66.67	59.79
Straight Hair	70.31	50.57	65.25	41.29
Receding Hairline	80.15	83.5	78.99	88.56
Bangs	81.98	0.84	100.0	0.42
Sideburns	77.71	57.31	68.06	49.49
Fully Visible Forehead	70.84	81.37	71.34	94.67
Partially Visible Forehead	93.97	0.0	0.0	0.0
Obstructed Forehead	85.04	0.0	0.0	0.0
Bushy Eyebrows	80.23	81.82	80.08	83.64
Arched Eyebrows	78.4	49.91	64.98	40.52
Narrow Eyes	72.98	80.36	75.42	85.99
Eyes Open	86.64	92.84	86.64	100.0
Big Nose	75.5	83.68	77.64	90.74
Pointy Nose	77.48	85.17	78.5	93.08
Big Lips	72.29	53.16	63.78	45.58
Mouth Closed	69.16	44.96	63.22	34.88
Mouth Slightly Open	72.29	50.48	69.81	39.53
Mouth Wide Open	90.69	0.0	0.0	0.0
Teeth Not Visible	71.07	75.05	70.02	80.85
No Beard	80.84	87.48	81.96	93.8
Goatee	80.23	47.03	75.16	34.23
Round Jaw	87.56	93.37	87.56	100.0
Double Chin	76.87	65.29	69.34	61.69
Wearing Hat	85.8	0.0	0.0	0.0
Oval Face	69.77	70.75	70.03	71.49
Square Face	95.04	0.0	0.0	0.0
Round Face	91.15	0.0	0.0	0.0
Color Photo	95.8	97.86	95.8	100.0
Posed Photo	63.74	64.79	63.89	65.71
Attractive Man	76.64	61.27	68.17	55.63
Attractive Woman	90.53	66.49	75.93	59.13
Indian	97.63	0.0	0.0	0.0
Gray Hair	86.26	32.33	61.43	21.94
Bags Under Eyes	70.76	77.13	72.83	81.98
Heavy Makeup	93.82	75.68	73.26	78.26
Rosy Cheeks	81.3	0.81	33.33	0.41
Shiny Skin	89.24	1.4	100.0	0.7
Pale Skin	61.68	64.35	60.56	68.64
5 o' Clock Shadow	75.65	69.24	71.66	66.98
Strong Nose-Mouth Lines	66.72	70.78	67.69	74.16
Wearing Lipstick	93.59	80.47	76.21	85.22
Flushed Face	86.79	0.0	0.0	0.0
High Cheekbones	74.81	58.12	75.33	47.31
Brown Eyes	72.6	80.41	77.42	83.65
Wearing Earrings	93.36	78.41	75.96	81.03
Wearing Necktie	78.02	83.56	78.54	89.27

Wearing Necklace	88.17	70.36	75.1	66.19