Your Age Revealed by Facebook Picture Metadata

Sanaz Eidizadehakhcheloo [1], **Bizhan Alipour Pijani** [2], Abdessamad Imine [2] & Michaël Rusinowitch [2]

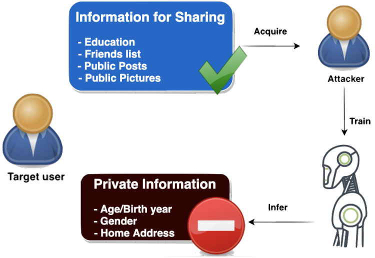
Sapienza University of Rome, Italy [1]
Lorraine University, Cnrs, Inria Nancy Grand Est, France [2]

July 13, 2020

1. Context
2. Motivation
3. Problem description
4. Existing works
5. Analysis process
6. Experiments
7. Conclusion

Inference attack

An Inference Attack is a data mining technique performed by analyzing data in order to illegitimately gain knowledge about a subject [**Wikipedia**].



Sources of inference attacks

- Behavioral records (liked pages, joined groups)
  - **You are how you behave**
- Based on your vicinity (friends list)
  - **Your friends shape who you are**
- Writing style (word and emoji usage)

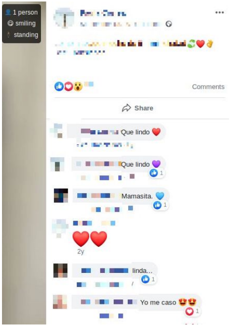
1. Context
2. **Motivation**
3. Problem description
4. Existing works
5. Analysis process
6. Experiments
7. Conclusion

Objectives of our work

- Raising awareness of Facebook users about their privacy
- Inference attacks based on seemingly innocent data

Picture alt-text, and comments

- Facebook attempted to involve blind people by launching automatic alt-text (AAT)
- People use comments to insert thoughts and feelings about the picture



1 Context

2 Motivation

3 Problem description

4 Existing works

5 Analysis process

6 Experiments

7 Conclusion

Problem description

Difficulty

Picture metadata have been added at different moments of the target lifetime and therefore relate to different age categories

Solution

Create different profiles with different age categories

1 Context

2 Motivation

3 Problem description

4 Existing works

5 Analysis process

6 Experiments

7 Conclusion

Existing works

Age inference attack on social media

- Private Traits and Attributes are Predictable from Digital Records of Human Behavior [**kosinski2013private**].
- Analyzing Biases in Human Perception of User Age and Gender from Text [**flekova2016analyzing**].
- Author Age Prediction from Text using Linear Regression [**nguyen2011author**].

Advantages over previous works

- Target user personality does not affect the performance
- The attack is still possible even when the entire personal data is unavailable
- The attack is suitable for online mode

Picture metadata

- Gender Inference for Facebook Picture Owners [**Alipour2019gender**].
- You are What Emojis Say About Your Pictures: Language-Independent Gender Inference Attack on Facebook [**Alipour2020you**].

1 Context

2 Motivation

3 Problem description

4 Existing works

5 Analysis process

6 Experiments

7 Conclusion

Data crawling and Pre-processing

Data crawling

- Collecting picture metadata, alt-text and comments from the HTML part of the picture
- Let U be the set of all users
- each user $u \in U$ a triple (age_u, A_u, C_u) where:
  - age_u is the age group of u
  - A_u is the alt text
  - C_u is comments from commenters

Pre-processing

- Picture distribution
  - We consider the picture publishing time to avoid assigning a picture to a wrong age group
- Cleansing comments
  - Applying lemmatization (loves, loving, lovely → love)
  - Changing shape of words (Heeeelloooo, love u)
  - Eliminating stop words (the, a, an , in)

Feature selection and Feature extraction

Spearman rank-order correlation coefficient

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n} \ (-1 \leq r_s \leq 1)$$

- d_i is the difference between two ranked variables
- n is the age interval size (in our case *32* as the age interval is $[18, 50]$)

To be significant, the observed *spearman* value must be greater than the critical value (two-side $r_s \ll 0.001$, degree of freedom is *32*)

Spearman rank-order correlation coefficient advantages

- It is a non-parametric technique
- It is insensitive to outliers
- It can be applied to any sample size

Spearman rank-order correlation coefficient result

(a)

Responses	*spearman's rank correlation coefficient*
Wow reaction 😮	-0.806
Received comments	-0.756
Emoji usage in comments	-0.746
Love reaction ❤️	-0.686

(b)

Alt-text	*spearman's rank correlation coefficient*
closeup	-0.791
1 person	-0.750
2 people	-0.740
smiling	-0.723
2 people smiling	-0.720
selfie	-0.706
outdoor	-0.693
ocean	-0.640
child	0.625

(c)

Comments	*spearman's rank correlation coefficient*
pretty	-0.795
👌	-0.791
❤️	-0.789
😍	–0.786
🤗	-0.780
😘	-0.773
gorgeous	-0.729
love	-0.712
nice	-0.678
beautiful	-0.636

(d)

Comments	*spearman's rank correlation coefficient*
selfie ❤️	-0.829
1 person 😍	-0.819
selfie 😍	-0.817
closeup 😍	-0.792
selfie 💕	-0.794
1person 💕	-0.780
1 person cute	-0.780
1 person beautiful	-0.770
selfie beautiful	-0.779
1 person 😘	-0.769

Feature selection

By patterns

- Non-textual
  - Single emoji
  - Repeated emoji
- Textual
  - Single emoji
  - single repeated emoji
  - Several repeated/non-repeated emojis

By linguistic

Word, emoji, phrase, and alt-text tags occurrences (*n-grams* in a window of size $n \in 1,2,3$)

By correlational analysis

Co-occurrence network of alt-text and comments (containing emoji, word, *GIF*)

Feature extraction

LIME

A technique that explains a classification by ranking features according to their contribution to predict the final result

Contribution?	Feature	Value
+1.111	<BIAS>	1.000
+0.248	ulook	1.000
+0.213	u0001f602u0001f602	1.000
+0.169	youre	1.000
+0.161	loveit	0.000
+0.153	u0001f495	1.000
+0.073	sopretty	0.000
+0.070	lovethe	0.000
+0.069	u2764	0.000
+0.057	hahaha	1.000
+0.048	goodpicture	0.000
+0.045	greatpicture	0.000
+0.026	sohappy	0.000
+0.024	u0001f923u0001f923	0.000
+0.019	youhave	1.000
+0.018	socute	0.000

Univariate feature selection

Selects features that have a statistically significant relationship to the final result

1 Context

2 Motivation

3 Problem description

4 Existing works

5 Analysis process

6 Experiments

7 Conclusion

Dataset and Machine Learnings

Dataset

- *8,922* pictures; *6184* pictures received complete metadata, *1762* pictures received no comments.
- Facebook was unable to generate alt-text for *976* pictures.

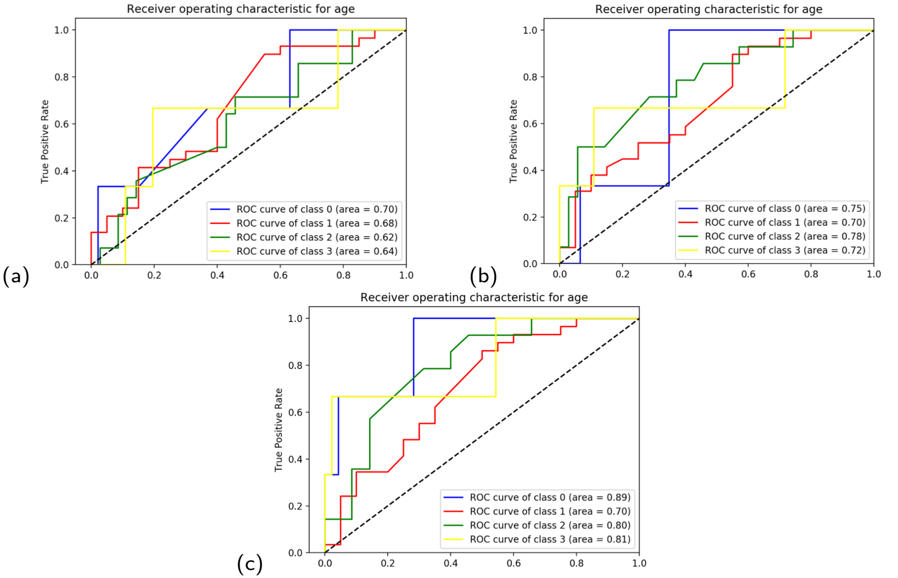
Machine Learning

- Logistic Regression classifier by using python library *scikit-learn*
- *80%* of it to train the classifier and *20%* to testing
- *5-fold* cross-validation on our training dataset
- *AUC-ROC* curve for evaluating the learning model

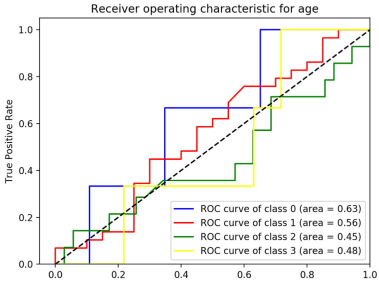
Age categories

- Class 0 → *18* to *20*
- Class 1 → *20* to *30*
- Class 2 → *30* to *40*
- Class 3 → *40* to *50*

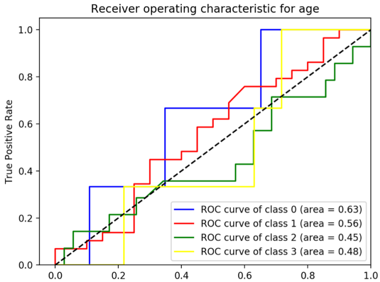
Age identification with best features



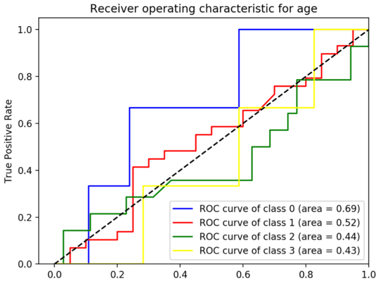
Age identification without best features



(d)



(e)



(f)

1. Context
2. Motivation
3. Problem description
4. Existing works
5. Analysis process
6. Experiments
7. Conclusion

Wrap up

Conclusion

- We identify that commenters react differently to younger and older owner pictures
- We have shown the possibility of age inference attack from limited available data

Advantages of this work

- Exploits data that are easily available
- Image processing is needless
- Words, emojis, and GIF handle all possible received comments

Perspectives

- Automatically extract features by using a *Random Walk* approach
- Adjust the age groups to infer the owner exact age
- Enrich the dataset to compare the result of deep learning approaches

Questions