User attribute prediction from social media posts

Daniel Preoţiuc-Pietro

Positive Psychology Center
University of Pennsylvania

Penn | World Well-Being Project

20 April 2016

User Profiling

Hypothesis

Online behaviors reveal individual differences in both demographic and psychological traits.

Types of behavior:

- **Text**
- **Profile Picture**
- Social Network
- Like/Share
- Check-ins

Motivation

User attribute prediction is successful.

Demographic traits:

- Age (Rao et al. 2010 ACL)
- Gender (Burger et al. 2011 EMNLP)
- Location (Eisenstein et al. 2010 EMNLP)
- Popularity (Lampos et al. 2014 EACL)
- Political Orientation (Volkova et al. 2014 ACL)

Psychological traits:

- Personality (Schwartz et al. 2013 PLoS One)
- Mental Illnesses (Coppersmith et al. 2014 ACL)

Motivation

User attribute prediction from text is useful:

- Social science research
- Large-scale population studies
- Real-time measurement
- Audience profiling and targeted marketing (Yang et al. 2015 EMNLP)
- Improving Natural Language Processing tasks: sentiment analysis (Volkova et al. 2013 EMNLP), text classification (Hovy 2015 ACL)

Motivation

Two sides of the problem:

Measurement using Machine Learning in a predictive setup (classification or regression)

Insight using interpretable features we can leverage this information to gain a better understanding of group differences

Motivation

Usual steps for user profiling:

- Data Acquisition
- Feature extraction and engineering
- Building a Machine Learning model
- Validate the model on a held-out set
- Interpret the results

Socio-Economic Factors

Socio-economic factors (occupation, social class, education, income) play a vital role in language use (Bernstein 1960, Labov 1972/2006)

Main contributions:

- Predicting new user attributes: occupation and income
- New dataset: user $\longleftrightarrow$ occupation $\longleftrightarrow$ income
- Gaussian Process classification for NLP tasks
- Feature ranking and analysis using non-linear methods

Standard Occupational Classification

Standardised job classification taxonomy

Developed and used by the UK Office for National Statistics (ONS), but applicable to any country

Hierarchical:

- 1-digit (major) groups: 9
- 2-digit (sub-major) groups: 25
- 3-digit (minor) groups: 90
- 4-digit (unit) groups: 369

Jobs grouped by **skill requirements**

Standard Occupational Classification

C1 Managers, Directors and Senior Officials

- 11 Corporate Managers and Directors
  - 111 Chief Executives and Senior Officials (£111,413)
    - 1115 Chief Executives and Senior Officials
      Job: chief executive, bank manager
    - 1116 Elected Officers and Representatives
  - 112 Production Managers and Directors (£50,952)
  - 113 Functional Managers and Directors (£70,457)
  - 115 Financial Institution Managers and Directors (£73,911)
  - 116 Managers and Directors in Transport and Logistics (£35,589)
  - 117 Senior Officers in Protective Services (£111,413)
  - 118 Health and Social Services Managers and Directors (£46,629)
  - 119 Managers and Directors in Retail and Wholesale (£29,009)
- 12 Other Managers and Proprietors

Standard Occupational Classification

C2 Professional Occupations
Job: mechanical engineer, pediatrist, postdoctoral researcher

C3 Associate Professional and Technical Occupations
Job: system administrator, dispensing optician

C4 Administrative and Secretarial Occupations
Job: legal clerk, company secretary

C5 Skilled Trades Occupations
Job: electrical fitter, tailor

C6 Caring, Leisure, Other Service Occupations
Job: school assistant, hairdresser

C7 Sales and Customer Service Occupations
Job: sales assistant, telephonist

C8 Process, Plant and Machine Operatives
Job: factory worker, van driver

C9 Elementary Occupations
Job: shelf stacker, bartender

Data

5191 users $\longleftrightarrow$ 3-digit job group $\longleftrightarrow$ mean income

Users collected by self-disclosure of job title in profile

Manually filtered by the authors

10M tweets, average 94.4 users per 3-digit group

Here we classify only at the 1-digit top level group (9 classes)

Feature representation and labels available online

Features

User Level features (**18**), such as:

- number of:
  - followers
  - friends
  - listings
  - tweets
- proportion of:
  - retweets
  - hashtags
  - @-replies
  - links
- average:
  - tweets/day
  - retweets/tweet

Features

Focus on **interpretable** features for analysis

Compute over reference corpus of 400M tweets:

- SVD embeddings and clusters
- Word2Vec(W2V) embeddings and cluster

SVD Features

Compute word × word similarity matrix

Similarity metric is Normalized PMI (Bouma 2009) using the entire tweet as context

SVD with different number of dimensions (30, 50, 100, 200)

User is represented by summing its word representations

The low-dimensional features offer no interpretability

SVD Features

Spectral clustering to get hard clusters of words (30, 50, 100, 200 clusters)

Each cluster consists of distributionally similar words $\longleftrightarrow$ topic

User is represented by the number of times he uses a word from each cluster

Word2Vec Features

Trained Word2Vec (layer size 50) on our Twitter reference corpus

Spectral clustering on the word $\times$ word similarity matrix (30, 50, 100, 200 clusters)

Similarity is cosine similarity of words in the embedding space

Gaussian Processes

Brings together several key ideas in one framework:

- Bayesian
- kernelised
- non-parametric
- non-linear

Elegant and powerful framework, with growing popularity in machine learning and application domains

Gaussian Process Classification

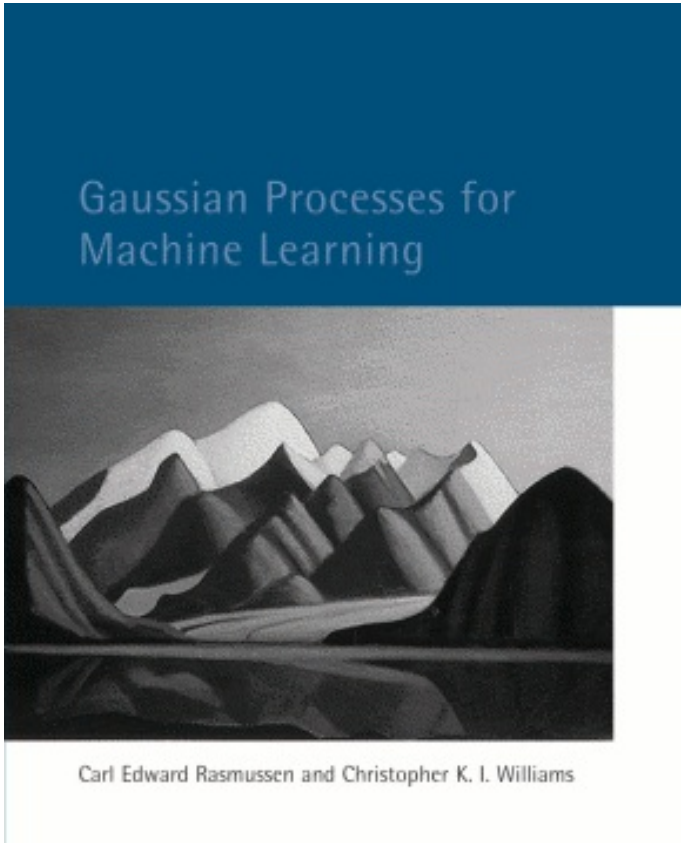
ARD kernel learns feature importance → features most **discriminative** between classes

We learn 9 one-vs-all binary classifiers

This way, we find the most predictive features consistent for all classes

Gaussian Process Resources

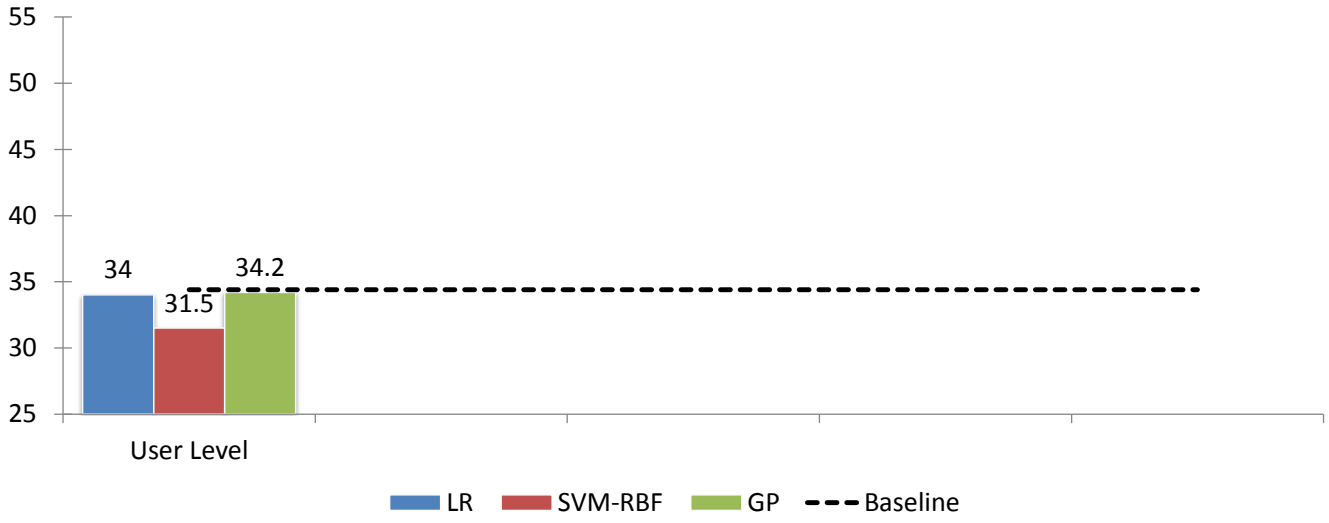
Free book:
`http://www.gaussianprocess.org/gpml/chapter`



Gaussian Process Resources

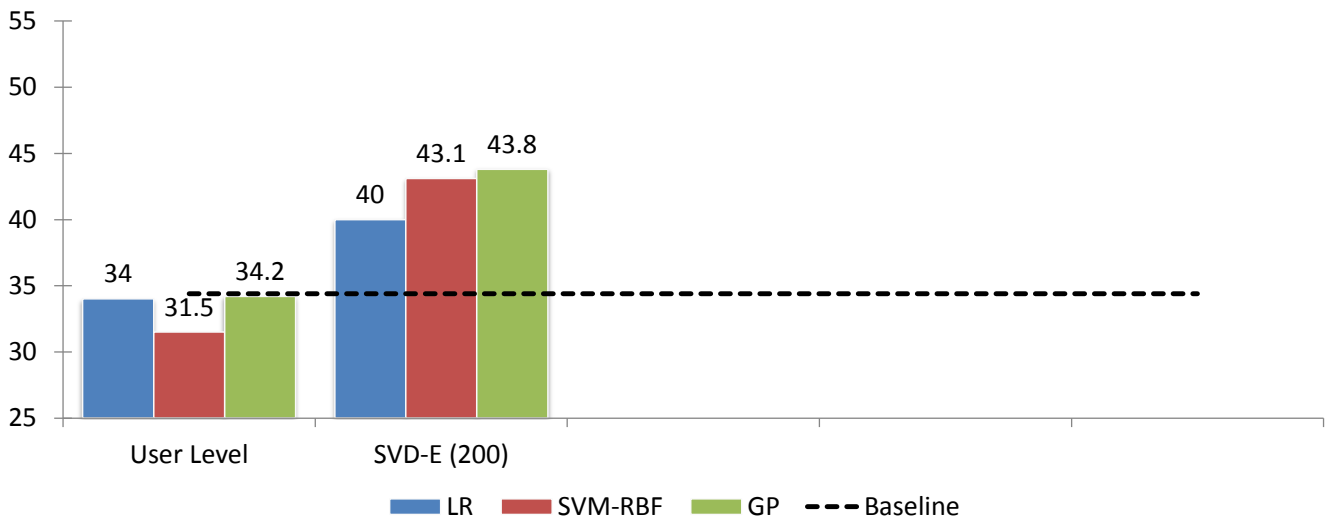
- GPs for Natural Language Processing tutorial (ACL 2014) `http://www.preotiuc.ro`
- GP School in Sheffield and roadshows in Kampala, Pereira, Nyeri, Melbourne `http://ml.dcs.shef.ac.uk/gpss/`
- Annotated bibliography and other materials `http://www.gaussianprocess.org`
- GPy Toolkit (Python) `https://github.com/SheffieldML/GPy`

Prediction



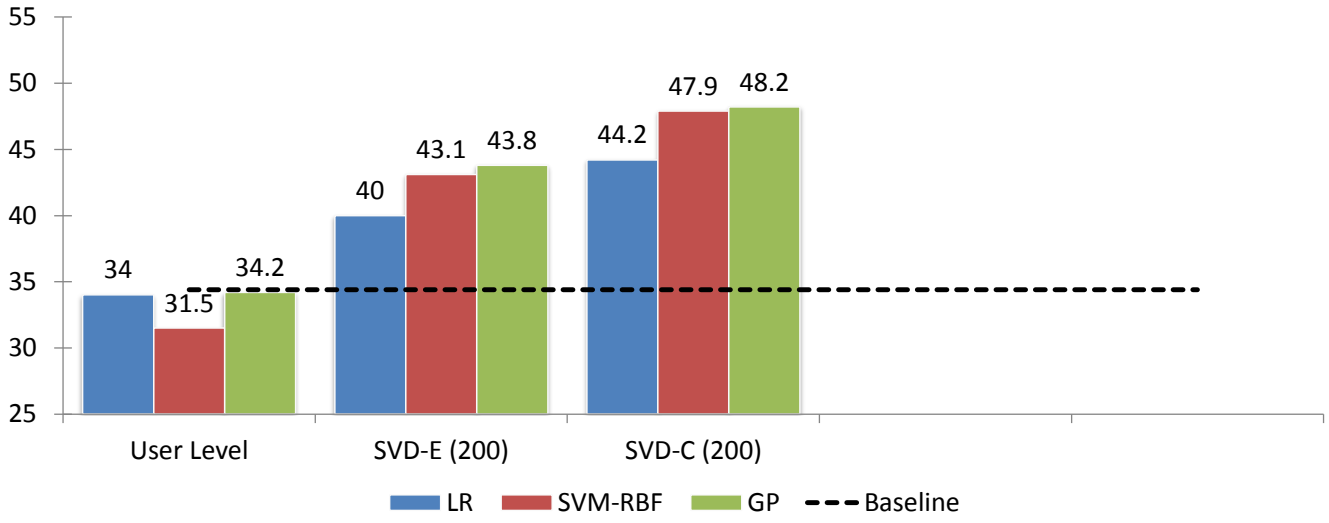
Stratified 10 fold cross-validation

Prediction



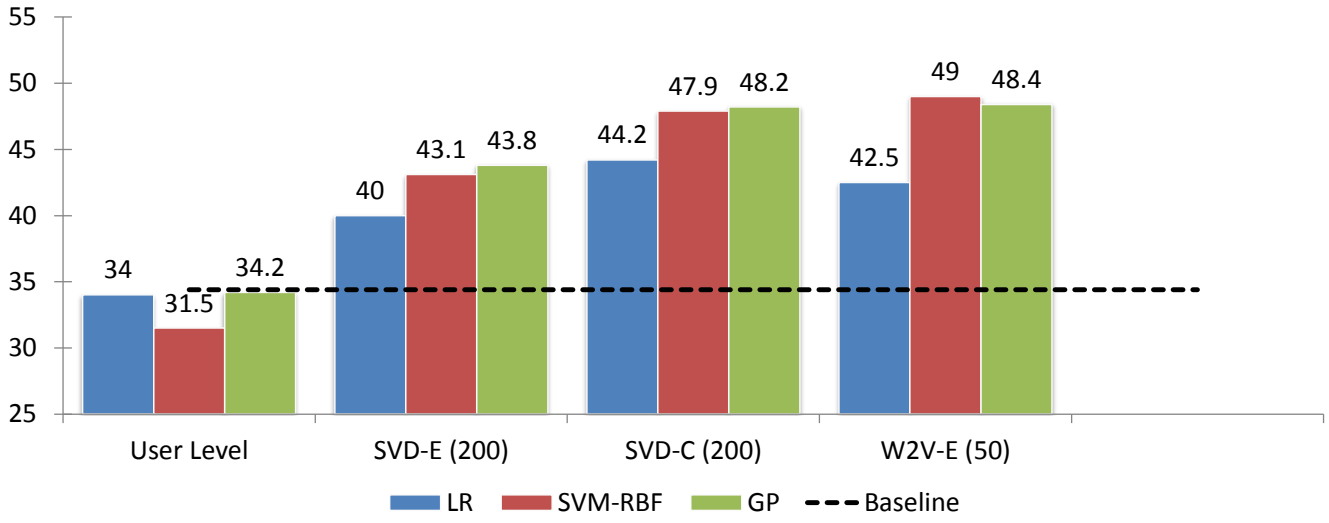
Stratified 10 fold cross-validation

Prediction



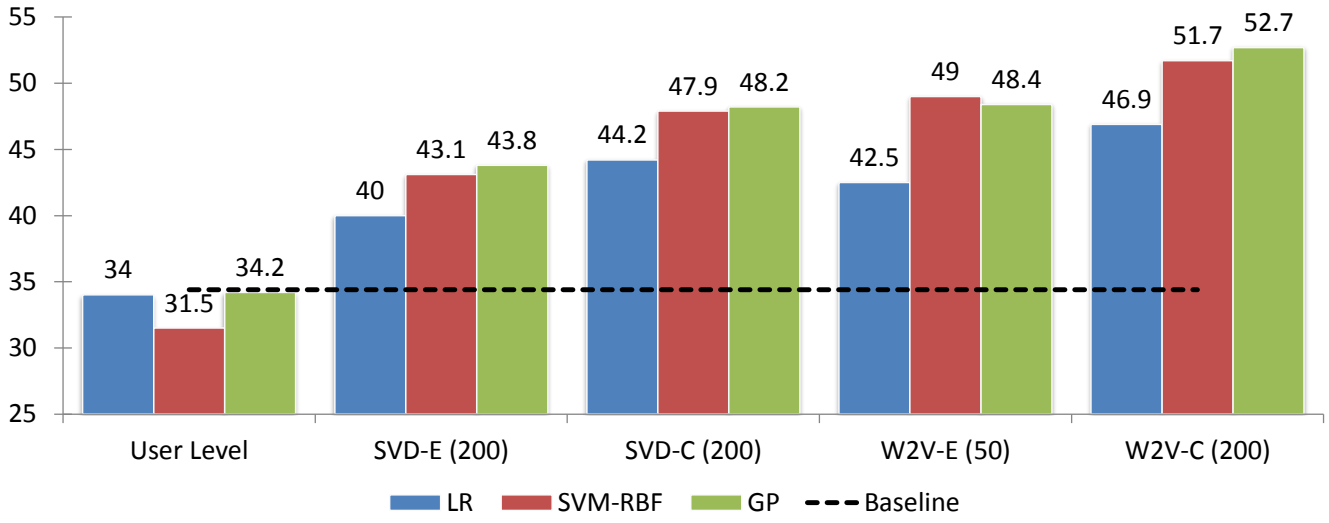
Stratified 10 fold cross-validation

Prediction



Stratified 10 fold cross-validation

Prediction



Stratified 10 fold cross-validation

Prediction Analysis

User level features have no predictive value

Clusters outperform embeddings

Word2Vec features are better than SVD/NPMI for prediction

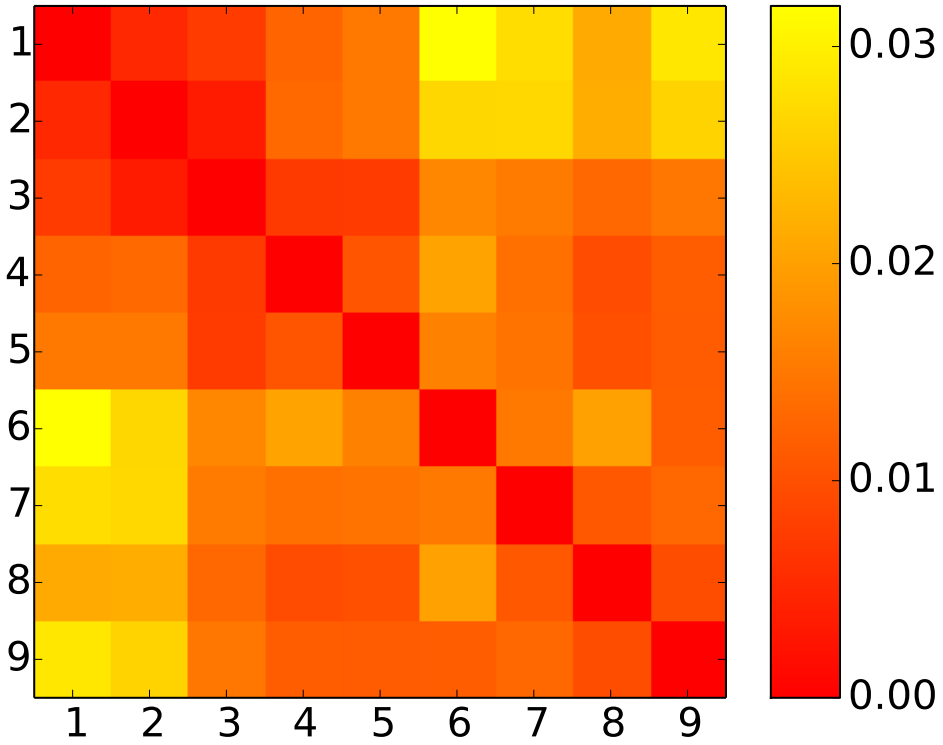
Non-linear methods (SVM-RBF and GP) significantly outperform linear methods

52.7% accuracy for 9-class classification is decent

Class Comparison

Jensen-Shannon Divergence between topic distributions across occupational classes

Some clusters of occupations are observable



Feature Analysis

Rank	**Manual Label**	**Topic** (most frequent words)
1	Arts	art, design, print, collection, poster, painting, custom, logo, printing, drawing
2	Health	risk, cancer, mental, stress, patients, treatment, surgery, disease, drugs, doctor
3	Beauty Care	beauty, natural, dry, skin, massage, plastic, spray, facial, treatments, soap
4	Higher Education	students, research, board, student, college, education, library, schools, teaching, teachers
5	Software Engineering	service, data, system, services, access, security, development, software, testing, standard

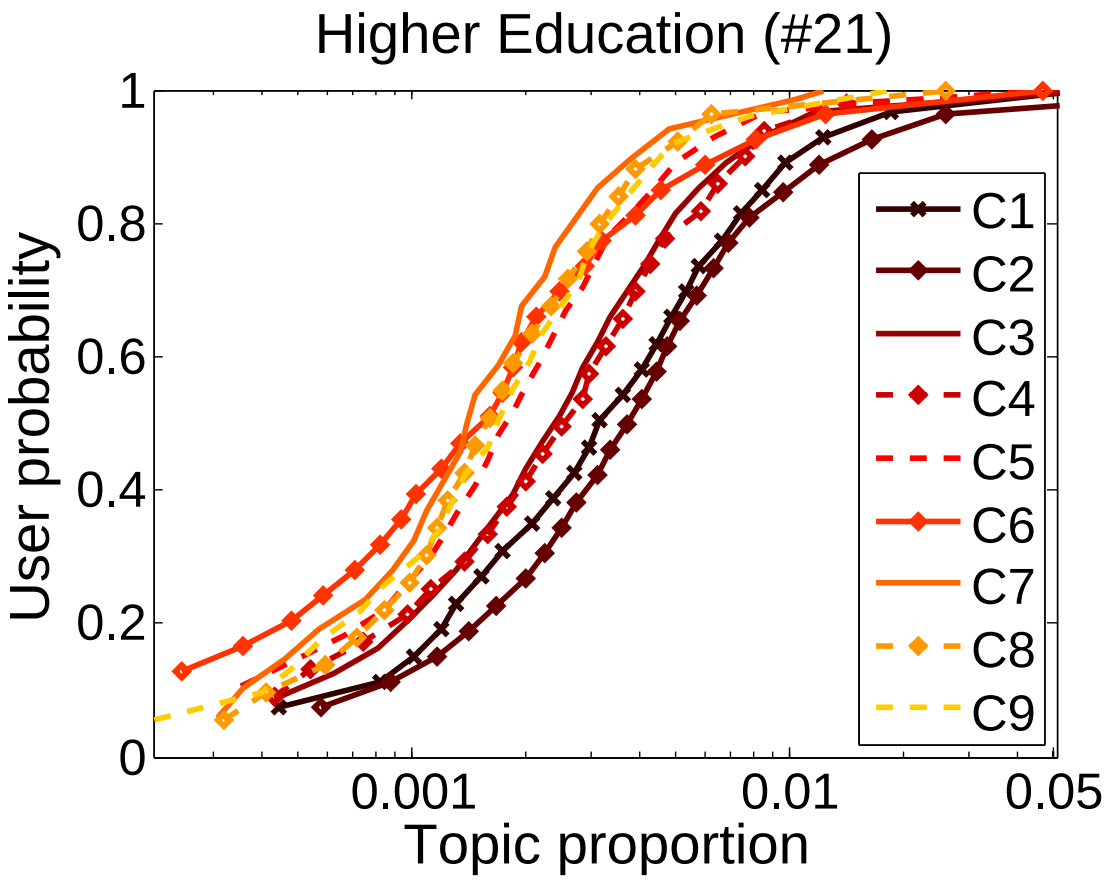
Most predictive Word2Vec 200 clusters as given by Gaussian Process ARD ranking

Feature Analysis

Rank	**Manual Label**	**Topic** (most frequent words)
7	Football	van, foster, cole, winger, terry, reckons, youngster, rooney, fielding, kenny
8	Corporate	patent, industry, reports, global, survey, leading, firm, 2015, innovation, financial
9	Cooking	recipe, meat, salad, egg, soup, sauce, beef, served, pork, rice
12	Elongated Words	wait, till, til, yay, ahhh, hoo, woo, woot, whoop, woohoo
16	Politics	human, culture, justice, religion, democracy, religious, humanity, tradition, ancient, racism

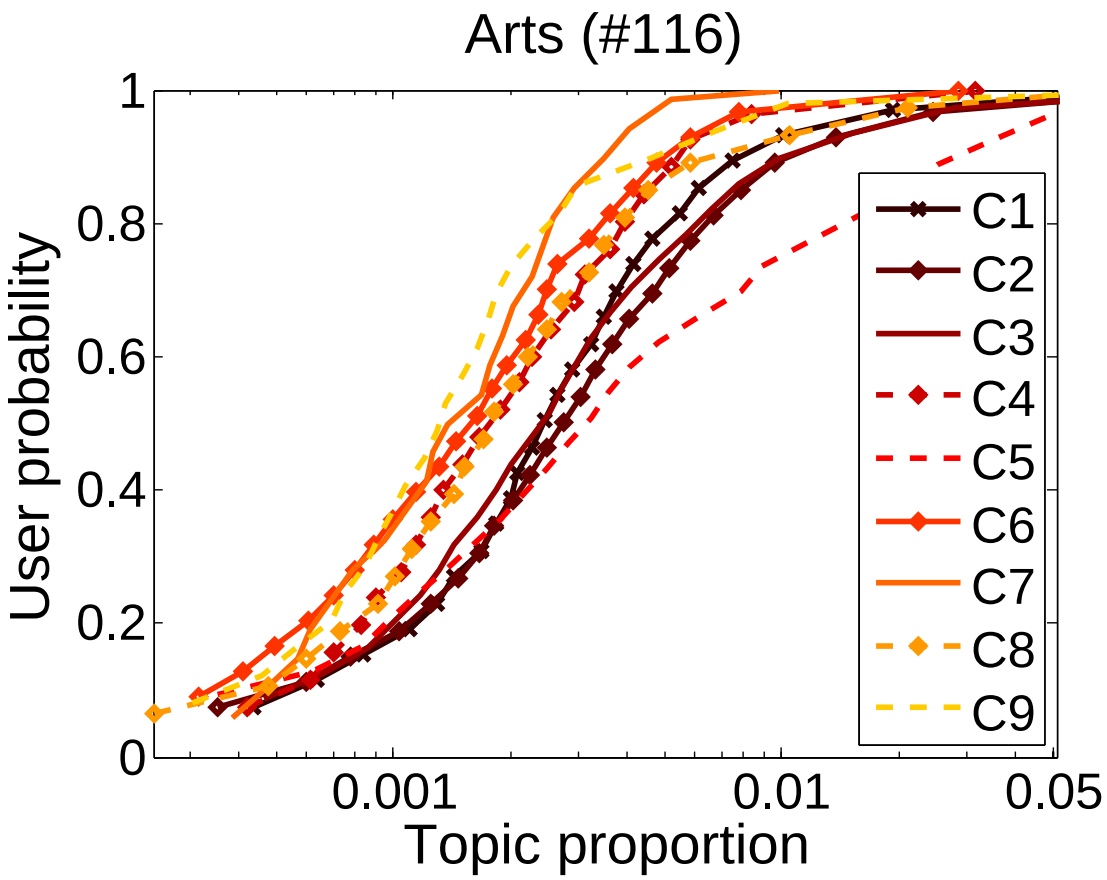
Most predictive Word2Vec 200 clusters as given by Gaussian Process ARD ranking

Feature Analysis - Cumulative Density Functions



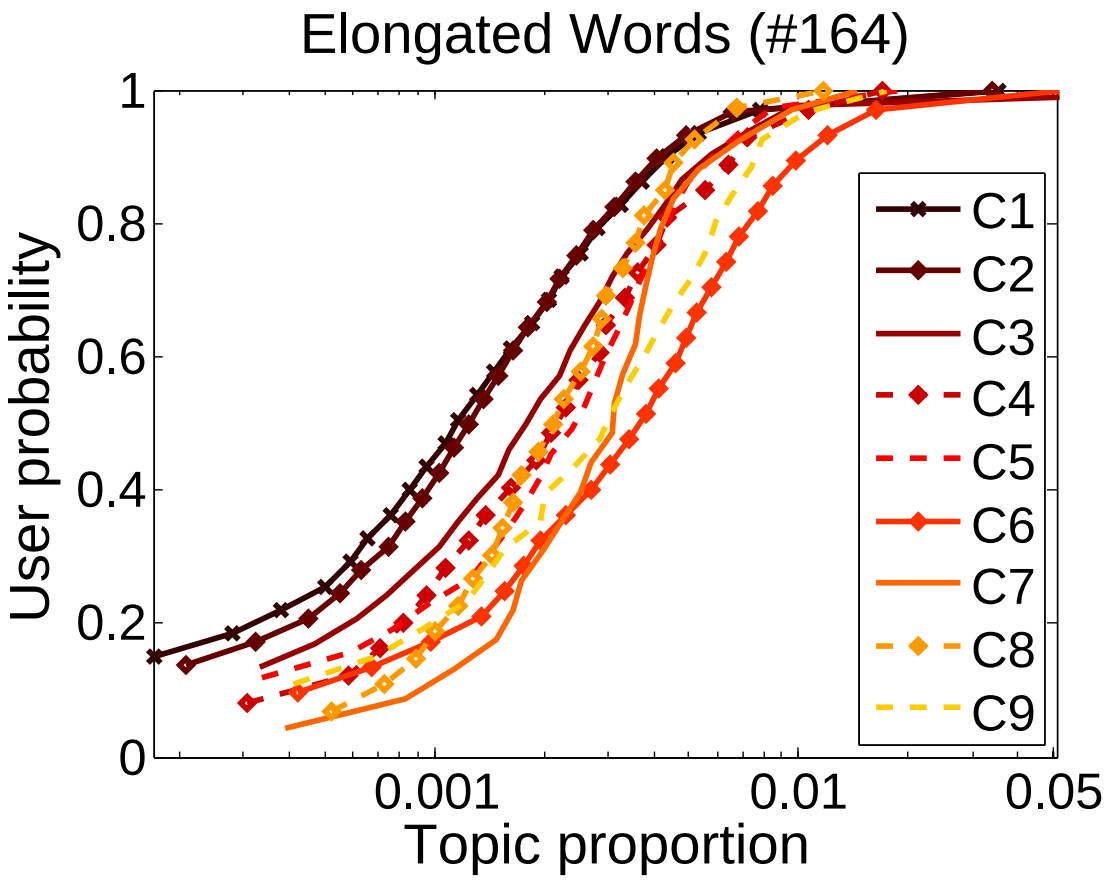
Topic more prevalent → CDF line closer to bottom-right corner

Feature Analysis - Cumulative Density Functions



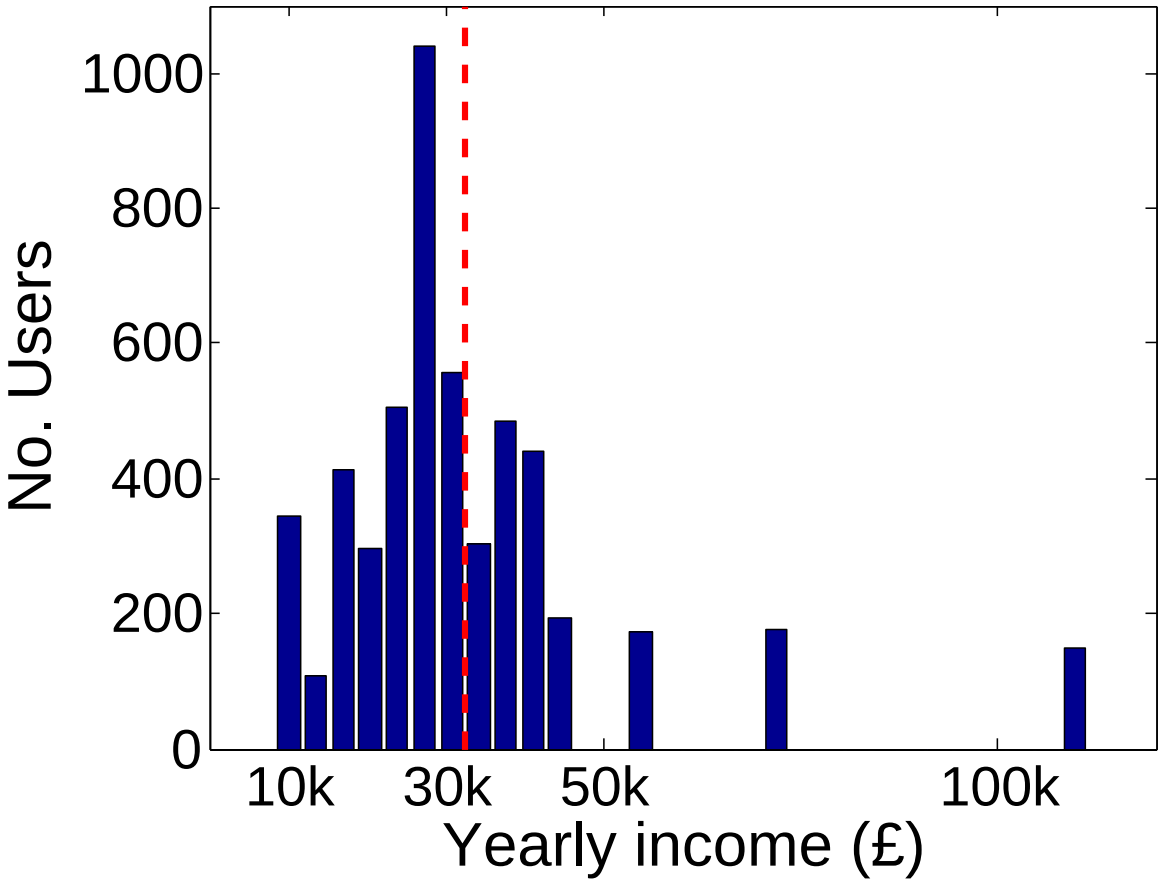
Topic more prevalent → CDF line closer to bottom-right corner

Feature Analysis - Cumulative Density Functions



Topic more prevalent → CDF line closer to bottom-right corner

Income prediction



We approach the task as regression.

Emotion Features

- Sentiment:
  - positive
  - neutral
  - negative
- Emotions:
  - anger
  - disgust
  - fear
  - joy
  - sadness
  - surprise

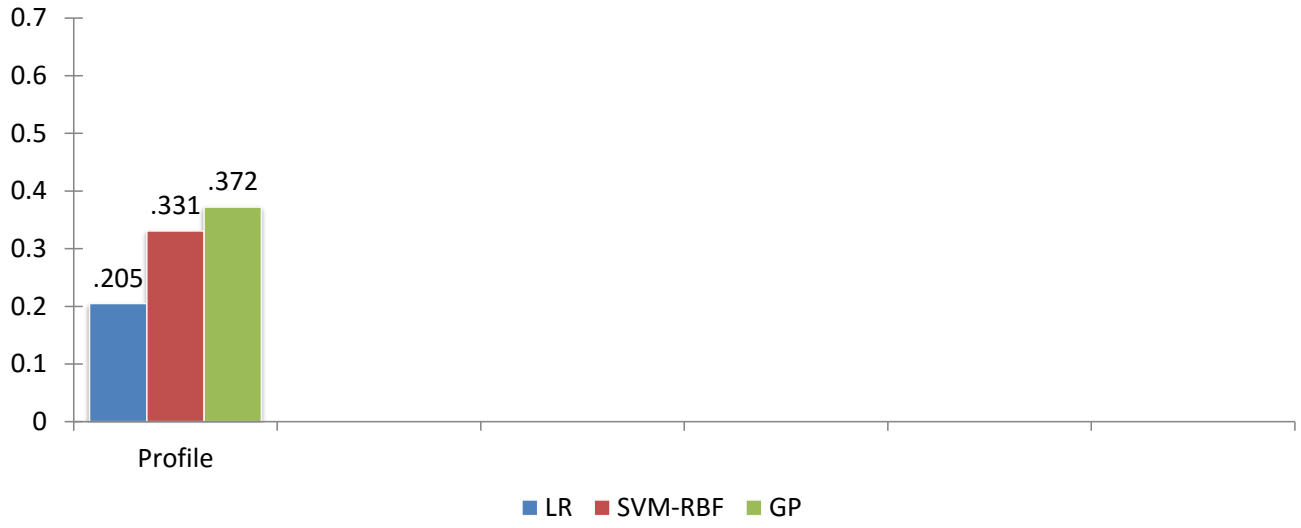
Predicted User Psycho-Demographic Features

- gender
  male, female
- age
- political
  independent, conservative, liberal, unaffiliated
- intelligence
  > average, average, < average, >> average, << average
- relationship
  married, in a relationship, single, other
- ethnicity
  Asian, African American, Indian, Hispanic, Other, Caucasian
- education
  bachelor, graduate, high school
- religion
  Christian, Jewish, Muslim, Hindu, unaffiliated, other
- children
  yes, no

Predicted User Psycho-Demographic Features

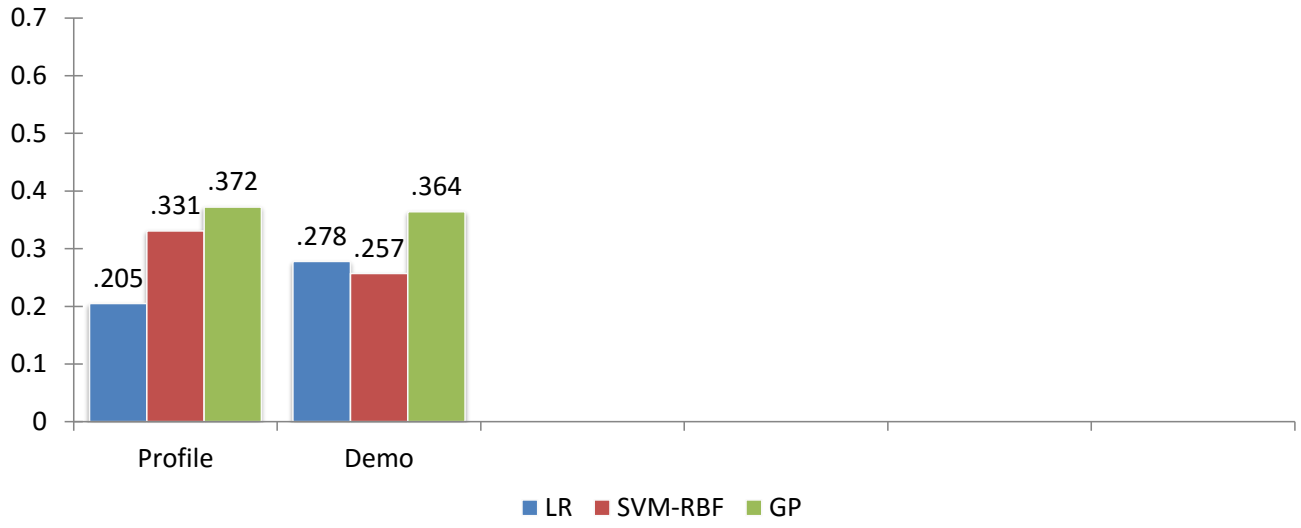
- income
  below average, above average, very high
- life satisfaction
  satisfied, dissatisfied, very satisfied, very dissatisfied, neither
- optimism
  optimist, pessimist, extreme optimist, extreme pessimist, neither
- narcissism
  agree strongly, agree, disagree, disagree strongly, neither
- excited
  agree strongly, agree, disagree, disagree strongly, neither
- anxious
  agree strongly, agree, disagree, disagree strongly, neither

Prediction



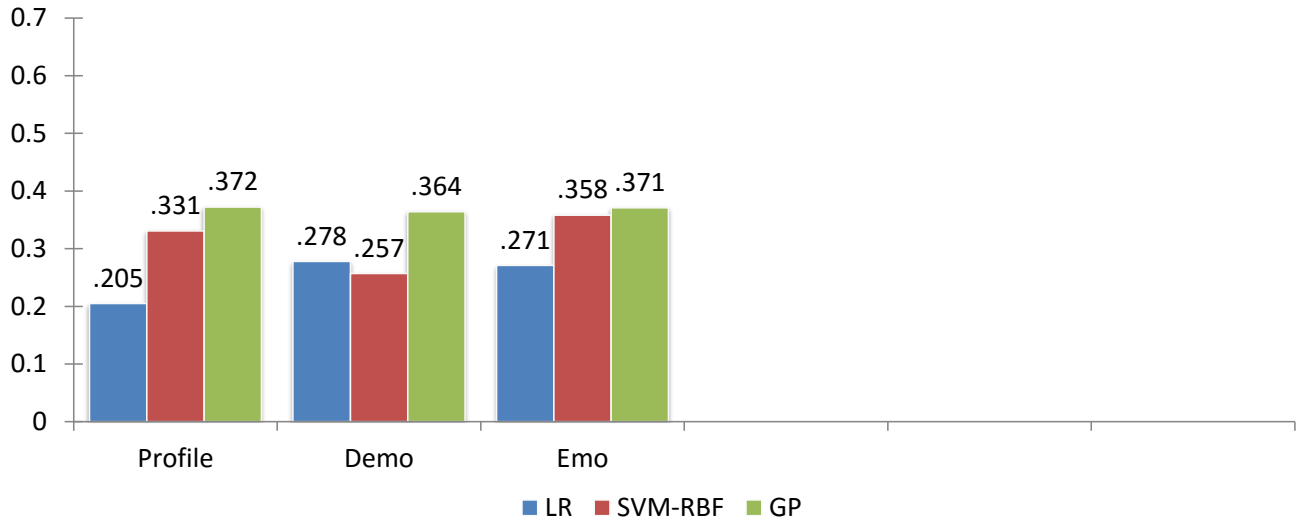
10 fold cross-validation

Prediction



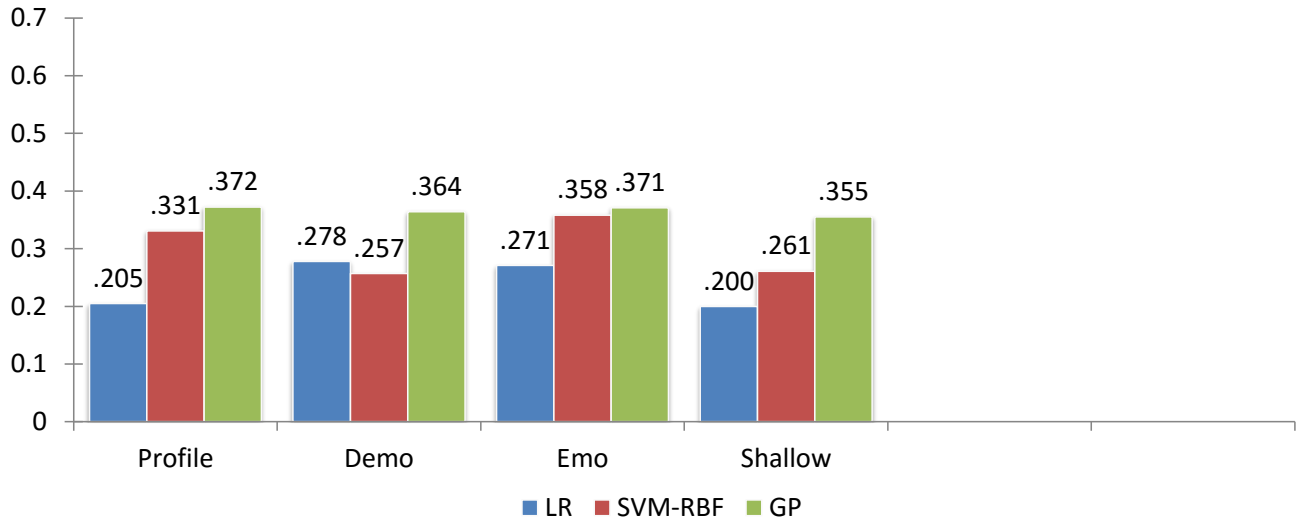
10 fold cross-validation

Prediction



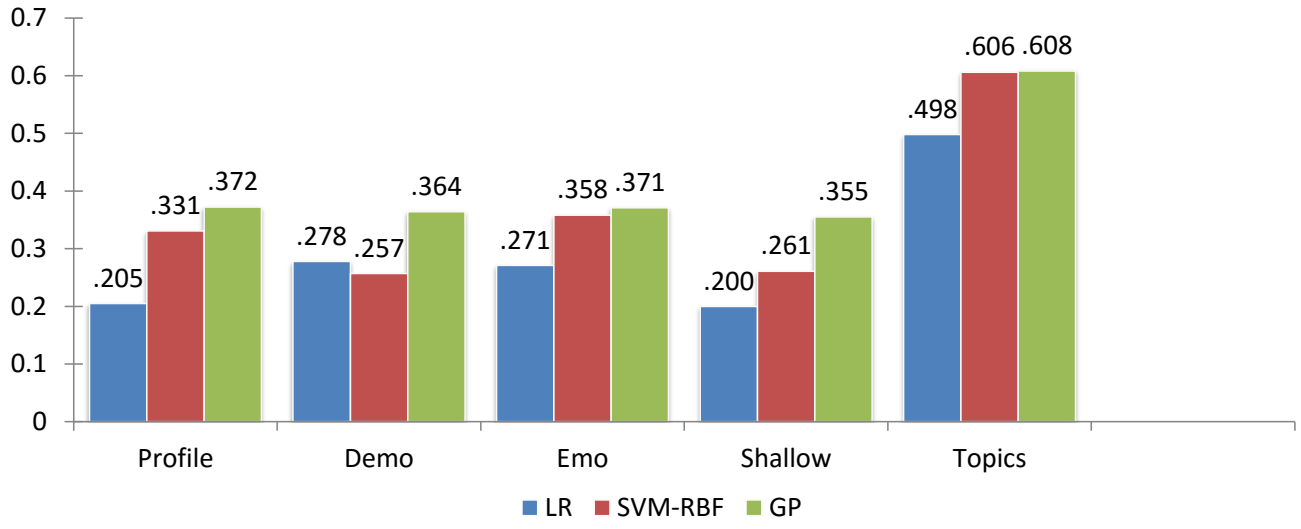
10 fold cross-validation

Prediction



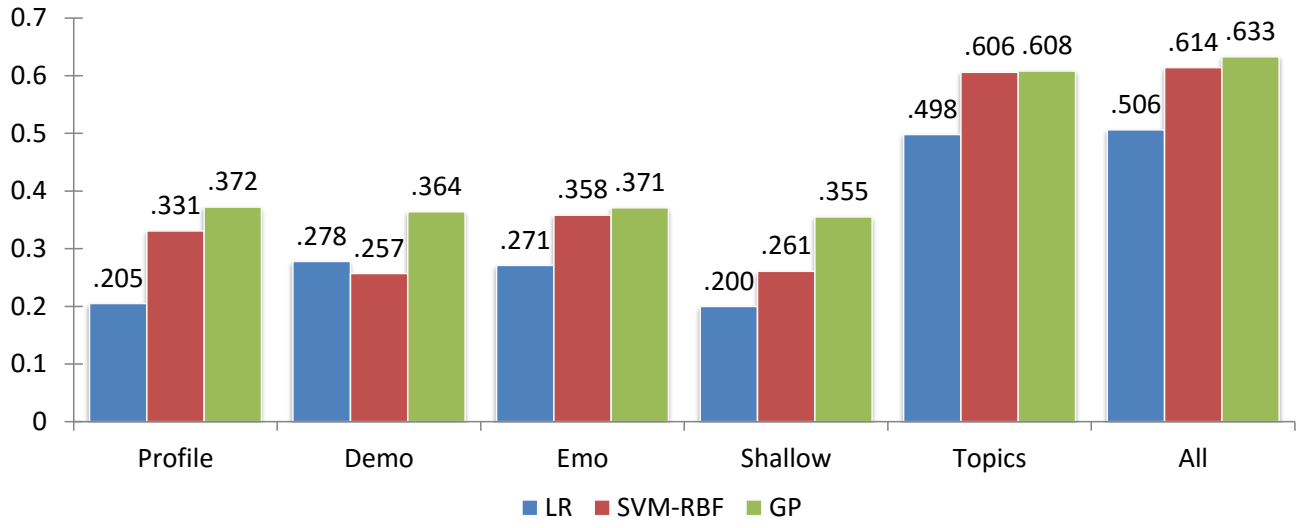
10 fold cross-validation

Prediction



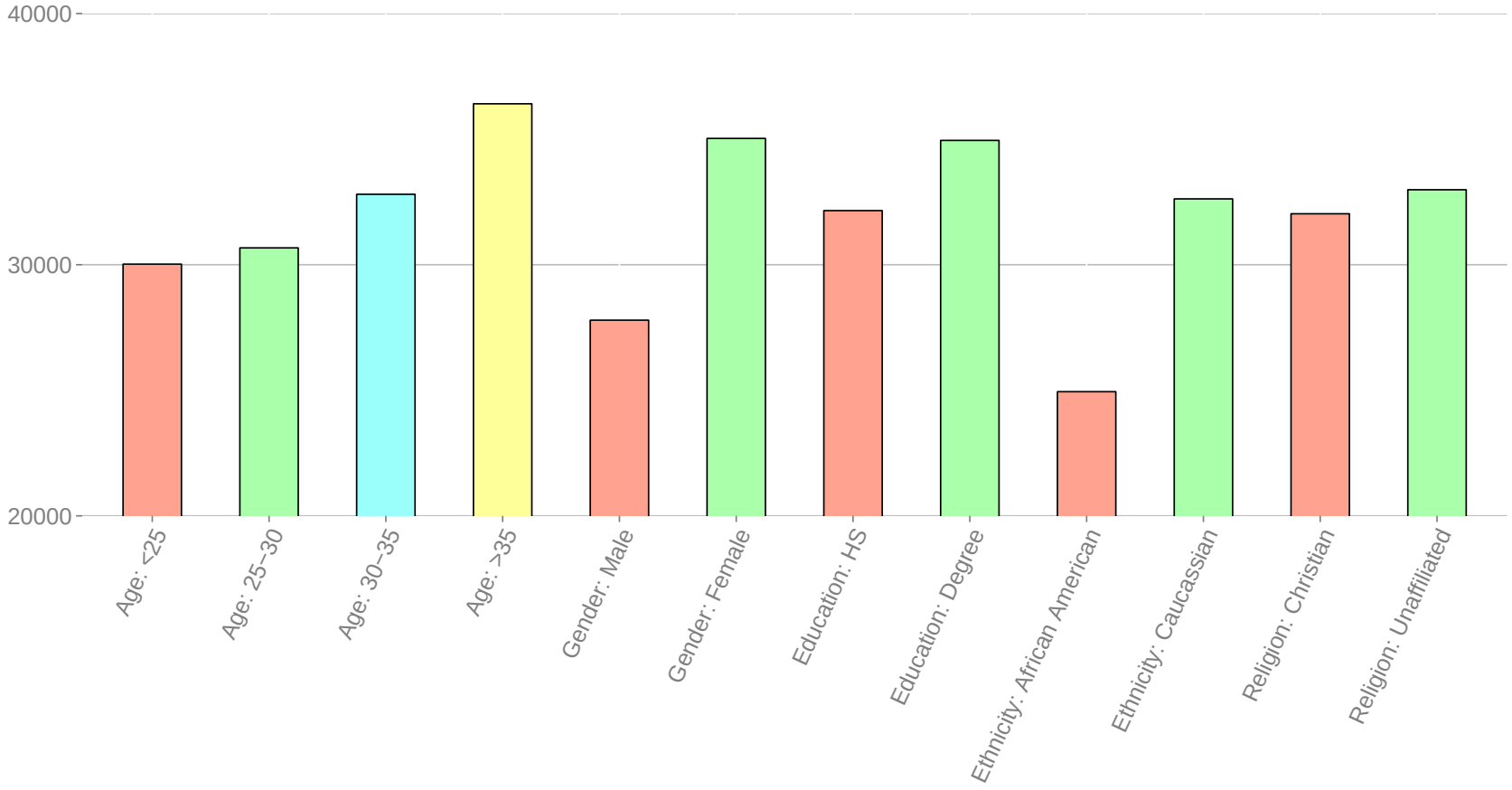
10 fold cross-validation

Prediction

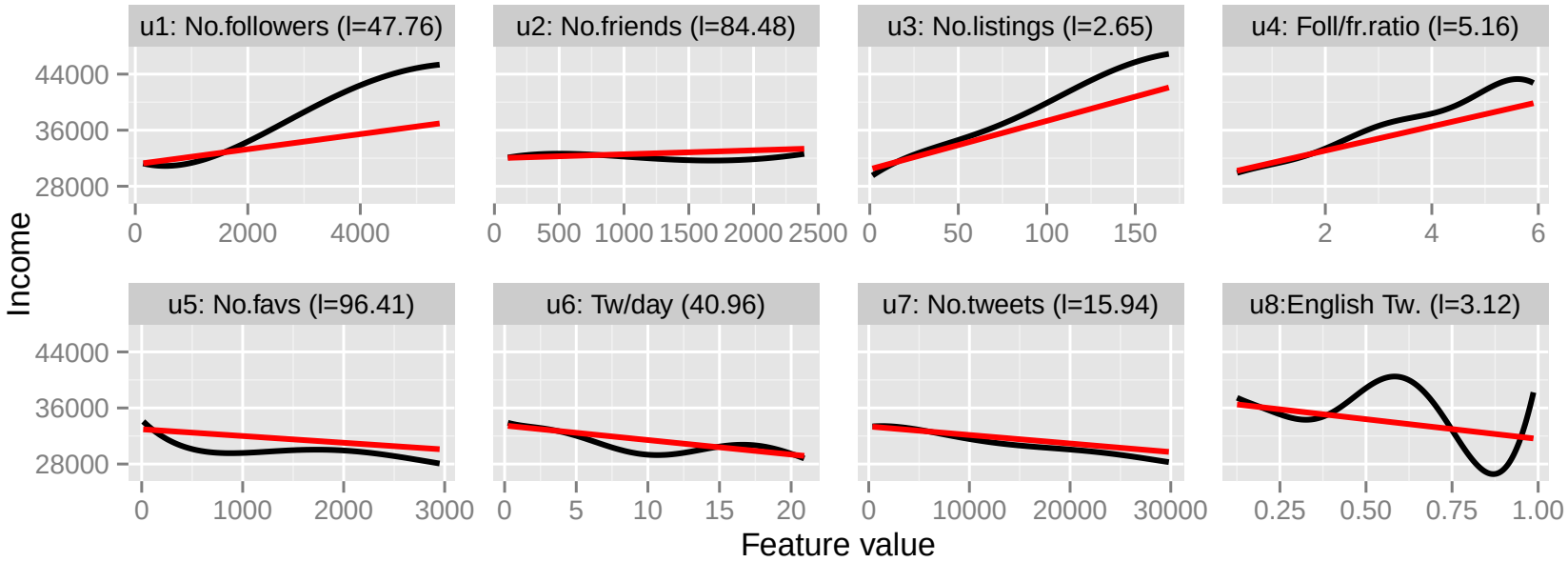


10 fold cross-validation

Psycho-Demographic Features

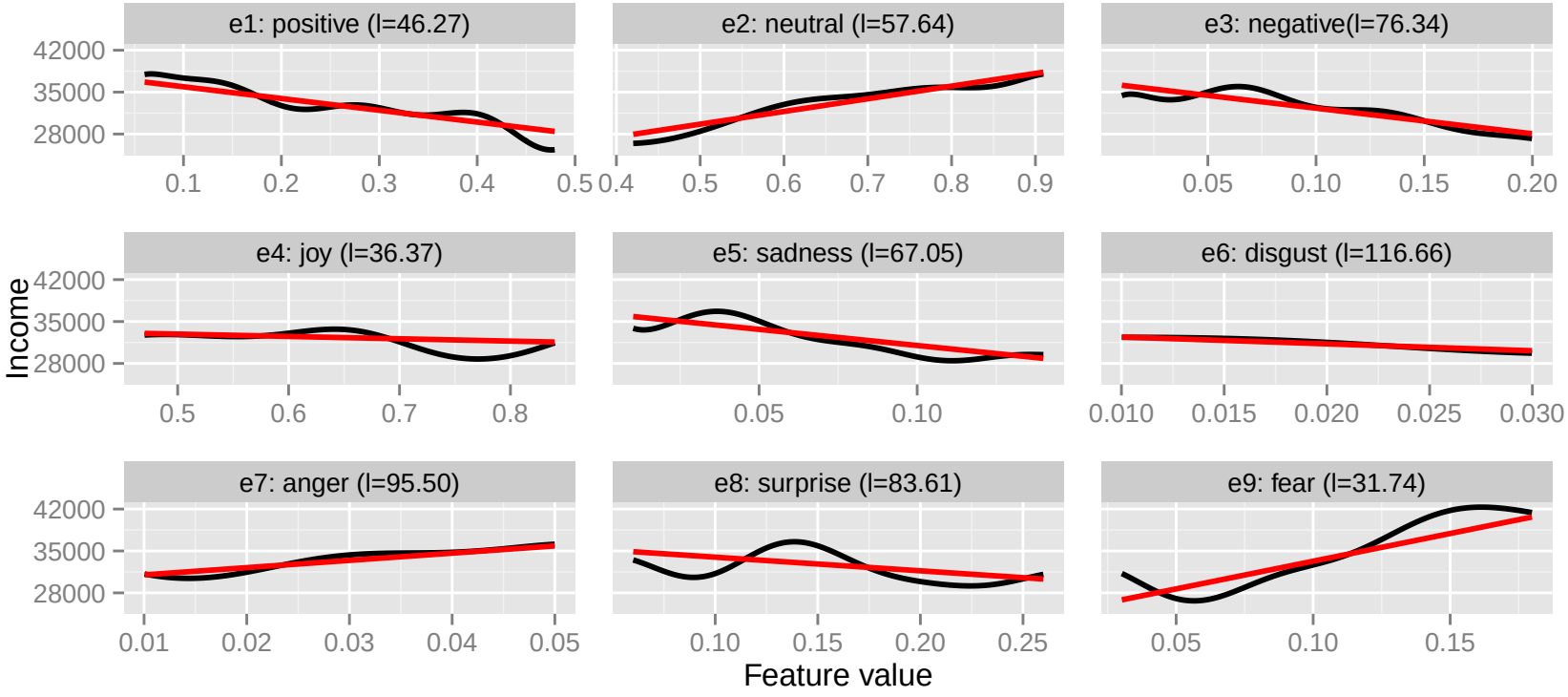


Profile Features



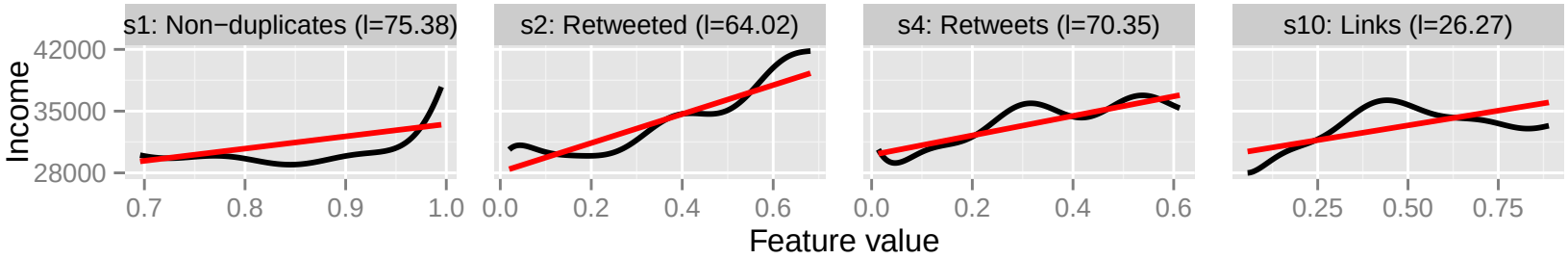
Linear fit, GP fit

Emotions



Linear fit, GP fit

Shallow Textual Features

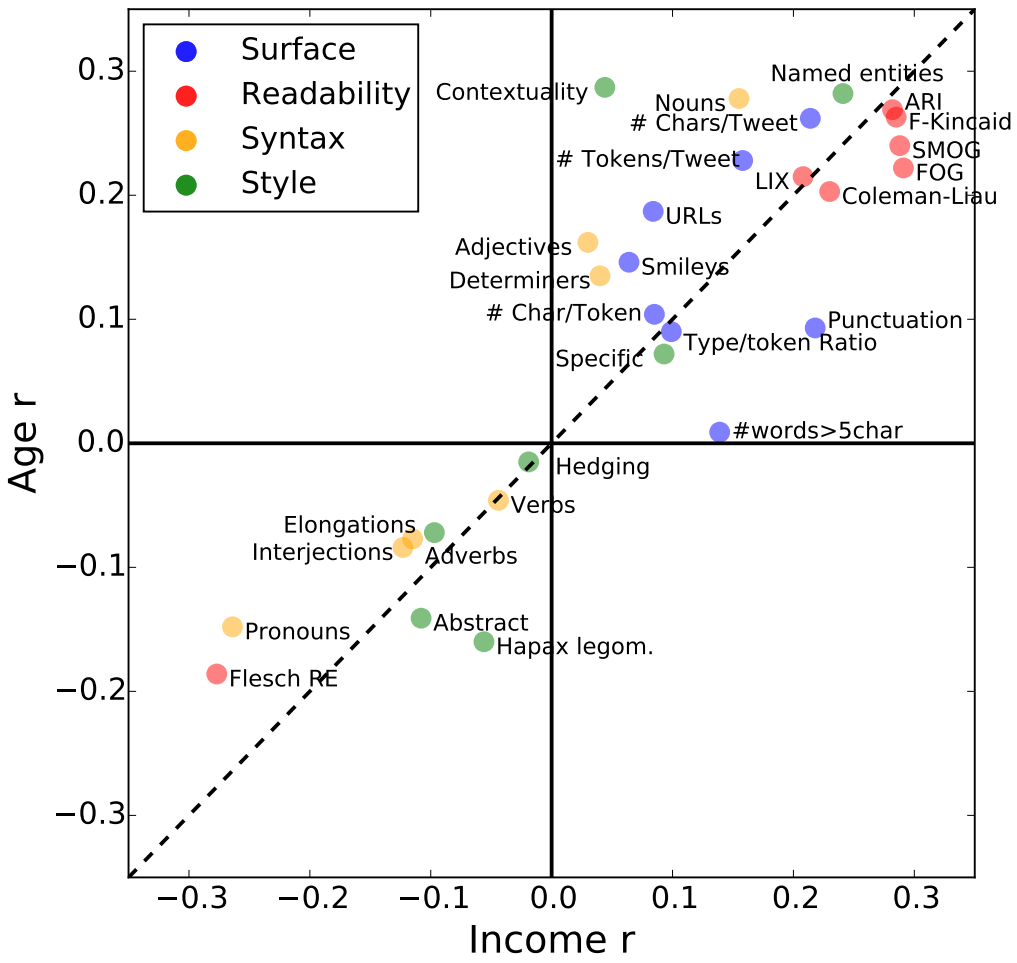


Linear fit, GP fit

Topics

Linear fit, GP fit

Stylistic changes with income and age



Take Aways

User socio-economic status influences language use in social media

Non-linear methods (Gaussian Processes) obtain significant gains over linear methods

Topic (clusters) features are both predictive and interpretable

New dataset available for research

Research Questions

1. **Can we automatically predict personality from profile picture choice?**

2. **What are the distinctive features of profile photos for each personality trait?**

Research Questions

1. Can we automatically predict personality from profile picture choice?

Yes! *(Celli et al. 2014), (Al Moubayed et al. 2014)*

2. What are the distinctive features of profile photos for each personality trait?

Bag-of-Visual-Words or Deep learning are hardly interpretable

Use facial and attractiveness features

Personality Guess

Which personality trait are users with these real Twitter Profile pictures high in?

Personality

- **Openness to experience**:
  - (**+**) inventive/curious – consistent/cautious (**-**)
- **Conscientiousness**:
  - (**+**) efficient/organized – easy-going/careless (**-**)
- **Extraversion**:
  - (**+**) outgoing/energetic – solitary/reserved (**-**)
- **Agreableness**:
  - (**+**) friendly/compassionate – analytical/detached (**-**)
- **Neuroticism**:
  - (**+**) sensitive/nervous – secure/confident (**-**)

Personality Guess

Which personality trait are users with these real Twitter Profile pictures high in?

Personality Guess

Which personality trait are users with these real Twitter Profile pictures high in?

+ Extraversion

+ Conscientiousness

Personality Guess

Twitter profile pictures - an image the user considers representative for their online persona

Personality prediction from standard photos is a relatively well studied problem in psychology (*Penton-Voak et al. 2006, Naumann et al. 2009*)

Humans are good at predicting some personality traits from a single photo: extraversion

Dataset

- 66,502 Twitter users
- text predicted personality
- self-reported gender
- text predicted age
- 104,500,740 tweets

Image Features - Color

- Image is Grayscale?
- RGB Spectrum
  - Red
  - Green
  - Blue
  - Average
- Brightness
- Contrast
- Saturation
- Hue
- Colorfulness
- Naturalness
- Sharpness
- Blur
- Color Emotions

Human judgements of the attractiveness of images are influenced by:

- color distributions *(Huang, Wang, and Wu 2006)*
- aesthetic principles related to color composition *(Datta et al. 2006)*

Image Features - Color

- Image is Grayscale?
- RGB Spectrum
  - Red
  - Green
  - Blue
  - Average
- Brightness
- Contrast
- Saturation
- Hue
- Colorfulness
- Naturalness
- Sharpness
- Blur
- Color Emotions

Black/White photos are more 'artistic'

Previous research showed that colors from images are related to psychologic traits *(Wexner 1954)*

- red – 'exciting-stimulating', 'protective-defending'
- green – 'calm-peaceful-serene'
- blue – 'secure-comfortable', 'calm-peaceful-serene'

Image Features - Color

- Image is Grayscale?
- RGB Spectrum
  - Red
  - Green
  - Blue
  - Average
- Brightness
- Contrast
- Saturation
- Hue
- Colorfulness
- Naturalness
- Sharpness
- Blur
- Color Emotions

High saturation indicates vividness and chromatic purity, which are more appealing to the human eye

Colourfulness = The difference against gray *(San Pedro and Siersdorfer 2009)*

Naturalness = The degree of correspondence between images and human perception *(Huang, Wang, and Wu 2006)*

Sharpness = Measures coarseness or the degree of detail contained in an image.

A proxy for the quality of the photographing gear and photographer *(Ke, Tang, and Jing 2006)*

Image Features - Color

- Image is Grayscale?
- RGB Spectrum
  - Red
  - Green
  - Blue
  - Average
- Brightness
- Contrast
- Saturation
- Hue
- Colorfulness
- Naturalness
- Sharpness
- Blur
- Color Emotions

Affective tone of colors *(Wei-ning, Ying-lin, and Sheng-ming 2006)*
Represented by 17 color histogram features

Correlations

Feature	**Demographics**		**Personality Trait**				
Color	Gender	Age	Ope	Con	Ext	Agr	Neu
Grayscale	-.050	-.014	.050	-.031	-.012		.014
Red	.026				-.041		
Green	-.021	.012			.021		.011
Blue	-.022				.045		
Average RGB	.030	.015		.025	.033	.019	
Brightness	.019			.030		.022	
Contrast			.014	.016		.017	-.011
Saturation	.046	.014	.013				
Hue		-.022	-.019	-.017	.024		.015
Colorfulness			-.014		.041	.030	-.034
Naturalness		.026	-.017	.014	-.028	.015	-.013
Sharpness	-.053		.028	-.026	.016	-.022	
Blur		.053	-.016	.036		.021	
Average Color Emotions	.020		-.020			.023	-.016

Pearson correlations between profile image and Big Five personality controlled for age and gender and with age and gender (coded as 1 – female, 0 – male) separately. Positive correlation is highlighted with green (paler green $p < .01$, deeper green $p < .001$, two-tailed t-test) and negative correlation with red (paler red $p < .01$, deeper red $p < .001$, two-tailed t-test).

Interpretation

Aesthetically pleasing images:

+ Brightness, **+** Contrast, **+** Sharpness, **+** Saturation, **-** Blur

Artistic images:

- Naturalness, **+** Grayscale

All correlated with **Ope** !

Interpretation

Correlated with **Ope** & **Neu**:

- Colorfulness, **-** Color Emotions

However, **Neu** no significant correlations with any of the aesthetically pleasing features.

Interpretation

Correlated with **Agr**:

Anti-correlated with all the aesthetically pleasing features.

Highest correlations with ‘Average Color Emotions’, ‘Colorfulness’ and ‘Brightness’

Ext shows similar, albeit lower correlations.

Image Features - Composition

- Rule of Thirds
- Edge Distribution
- Hue Count
- Visual Weight
- Static Lines
- Dynamic Lines

Edge Distribution = Spatial distribution of the high frequency edges of an image

In good quality photos, the edges are focused on the subject

The number of unique hues of a photo is another measure of simplicity

Good compositions have fewer objects, resulting in fewer distinct hues *(Ke, Tang, and Jing 2006)*.

Visual weight measures the clarity contrast between subject region and the whole image

The presence of lines in an image induces emotional effects *(Arnheim 2004)*

Correlations

Feature	**Demographics**		**Personality Trait**				
Image Composition	Gender	Age	Ope	Con	Ext	Agr	Neu
Average Rule of Thirds	.036	.052	-.029	-.022	.038	.036	-.036
Edge Distribution	-.038	.016	.046			-.051	.039
Hue Count		.026	-.016				
Visual Weight				-.017			
Static Lines	.056				.018	.019	
Dynamic Lines	.044		-.024			.033	

Pearson correlations between profile image and Big Five personality controlled for age and gender and with age and gender (coded as 1 – female, 0 – male) separately. Positive correlation is highlighted with green (paler green $p < .01$, deeper green $p < .001$, two-tailed t-test) and negative correlation with red (paler red $p < .01$, deeper red $p < .001$, two-tailed t-test).

Interpretation

Again, aesthetically pleasing features are **+** with **Ope** and **-** with **Agr**, and to a lesser extent **-** with **Ext**.

The number of dynamic lines (indicative of emotional content) is **-Ope** and **+Agr**.

Image Features - Type

- Default Image
- Is Not Face
- One Face
- Multiple Faces
- No. Faces

Detected using Face++ API

Correlations

Feature	**Demographics**		**Personality Trait**				
Image Type	Gender	Age	Ope	Con	Ext	Agr	Neu
Default Image			-.022		-.043	.015	-.023
Is Not Face	-.072	-.021	.061	-.121	-.108	-.070	.071
One Face	.054	.029	-.016	.102	.081	.046	-.057
Multiple Faces	.040	-.019	-.102	.043	.058	.053	-.032
No. Faces	.072		-.092	.106	.103	.078	-.067

Pearson correlations between profile image and Big Five personality controlled for age and gender and with age and gender (coded as 1 – female, 0 – male) separately. Positive correlation is highlighted with green (paler green $p < .01$, deeper green $p < .001$, two-tailed t-test) and negative correlation with red (paler red $p < .01$, deeper red $p < .001$, two-tailed t-test).

Interpretation

Two ‘clusters’:

1) **Ope** & **Neu**

Not default picture and preferably no face

Multiple faces strongest **- Ope**

No face strongest **- Neu**

2) **Con** & **Ext** & **Agr**

One or more faces: all **+**.

Con strongest correlated with single face and strongest anti-correlated with no face.

Image Features - Demographics

- Age
- Gender
- Race
  - Asian
  - Black
  - White

Detected using Face++ API

Correlations

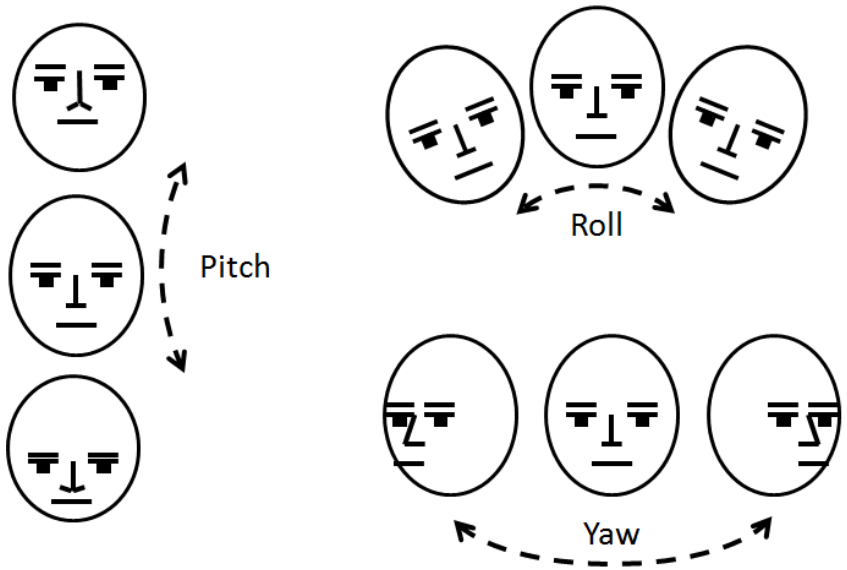
Feature	**Demographics**		**Personality Trait**				
Image Demographics	Gender	Age	Ope	Con	Ext	Agr	Neu
Age	-.310	.306	.050	.105	-.036		
Gender	.795	-.041			.035	.034	
Asian	.064	-.150	-.072	-.042			
Black	-.034	-.061	.047	.050	.085	-.055	-.096
White	-.033	.169	.031		-.066	.026	.071

Pearson correlations between profile image and Big Five personality controlled for age and gender and with age and gender (coded as 1 – female, 0 – male) separately. Positive correlation is highlighted with green (paler green $p < .01$, deeper green $p < .001$, two-tailed t-test) and negative correlation with red (paler red $p < .01$, deeper red $p < .001$, two-tailed t-test).

Image Features - Facial Presentation

- No Glasses
- Reading Glasses
- Sunglasses
- Pitch Angle
- Roll Angle
- Yaw Angle
- Face Ratio

Detected using Face++ API



Yaw – Usually predictive of selfies

Correlations

Feature	**Demographics**		**Personality Trait**				
Facial Presentation	Gender	Age	Ope	Con	Ext	Agr	Neu
No Glasses	.145	-.036		.027	.085	.026	-.065
Reading Glasses	-.141	.054	.020		-.099	-.017	.071
Sunglasses	-.034	-.020	-.017	-.028		-.019	
Pitch Angle	-.043						
Roll Angle	.017						
Yaw Angle							
Face Ratio	.034	.036	.038	-.039	-.097	-.039	.057

Pearson correlations between profile image and Big Five personality controlled for age and gender and with age and gender (coded as 1 – female, 0 – male) separately. Positive correlation is highlighted with green (paler green $p < .01$, deeper green $p < .001$, two-tailed t-test) and negative correlation with red (paler red $p < .01$, deeper red $p < .001$, two-tailed t-test).

Interpretation

Reading Glasses **+ Neu** and **- Ext, Agr**

Sunglasses **- Con**

Face ratio **+ Ope, Neu** and **- Con, Ext, Agr**

Combined with previous findings, **Ope & Neu** prefer no faces in picture, but when a face is present, this occupies a larger part of the photo.

Image Features - Facial Expression

- Smiling
- Anger
- Disgust
- Fear
- Joy
- Sadness
- Surprise
- Left Eye Openness
- Right Eye Openness
- Attention
- Expressiveness
- Neutral Expression4
- Positive Mood
- Negative Mood
- Valence

Smile detected using Face++ API

All other features detected using EmoVu

Expressiveness is the highest emotion value

Negative mood is the maximum value of the negative emotions (anger, disgust, fear, sadness)

Positive mood is the maximum value of the positive emotions (joy, surprise)

Valence is the average of positive and negative mood

Correlations

Feature	**Demographics**		**Personality Trait**				
Facial Expressions	Gender	Age	Ope	Con	Ext	Agr	Neu
Smiling	.229	.141	-.089	.190	.050	.148	-.104
Anger	-.108	-.019	.037	-.080	-.042	-.055	.056
Disgust	-.142	.048					
Fear		-.017	.018	-.029		-.043	.018
Joy	.191	.119	-.093	.180	.061	.140	-.107
Sadness	-.122	-.032	.023	-.051		-.034	.026
Surprise	.038	-.064		-.041		-.031	
Left Eye Openness	.093			.025			
Right Eye Openness	.091			.027			
Attention	-.055	.061	-.047	.049	.018	.040	-.048
Expressiveness	.101	.123	-.072	.140	.054	.106	-.089
Neutral	-.064	-.133	.068	-.128	-.047	-.093	.081
Positive Mood	.198	.111	-.093	.175	.065	.137	-.107
Negative Mood	-.164		.043	-.079	-.029	-.067	.044
Valence	.101	.132	-.075	.140	.053	.105	-.090

Pearson correlations between profile image and Big Five personality controlled for age and gender and with age and gender (coded as 1 – female, 0 – male) separately. Positive correlation is highlighted with green (paler green $p < .01$, deeper green $p < .001$, two-tailed t-test) and negative correlation with red (paler red $p < .01$, deeper red $p < .001$, two-tailed t-test).

Interpretation

Again, two ‘clusters’:

1) **Ope & Neu**

2) **Con & Ext & Agr**

Interpretation

1) **Ope & Neu**

- smiling

Emotions: **+** Anger, **+** Fear, **+** Sadness, **-** Joy

- Positive mood, **+** Negative mood, **-** Valence

- Attention, **-** Expressiveness, **+** Neutral

Interpretation

2) **Con & Ext & Agr**

Almost exact opposite of previous cluster.

Exceptions:

- Surprise cf. **0** Surprise

Eye Openness **+ Con**

Intriguingly, **Con** highest in all ‘positive’ emotions.

Ext lowest in all ‘positive’ emotions.

Overview - Openness

- artistic photos
- aesthetically pleasing
- low in color emotions
- less faces, especially more than one
- expressing more negative facial emotions
- less expressive, more neutral

Overview - Neuroticism

- neither artistic or not
- neither aesthetically pleasing or not
- low in color emotions
- less faces
- expressing strongest negative facial emotions
- less expressive, more neutral

Overview - Conscientiousness

- neither artistic or not
- neither aesthetically pleasing or not
- no relation with color emotions
- strongest preference for a single face
- expressing strongest positive facial emotions
- most expressive

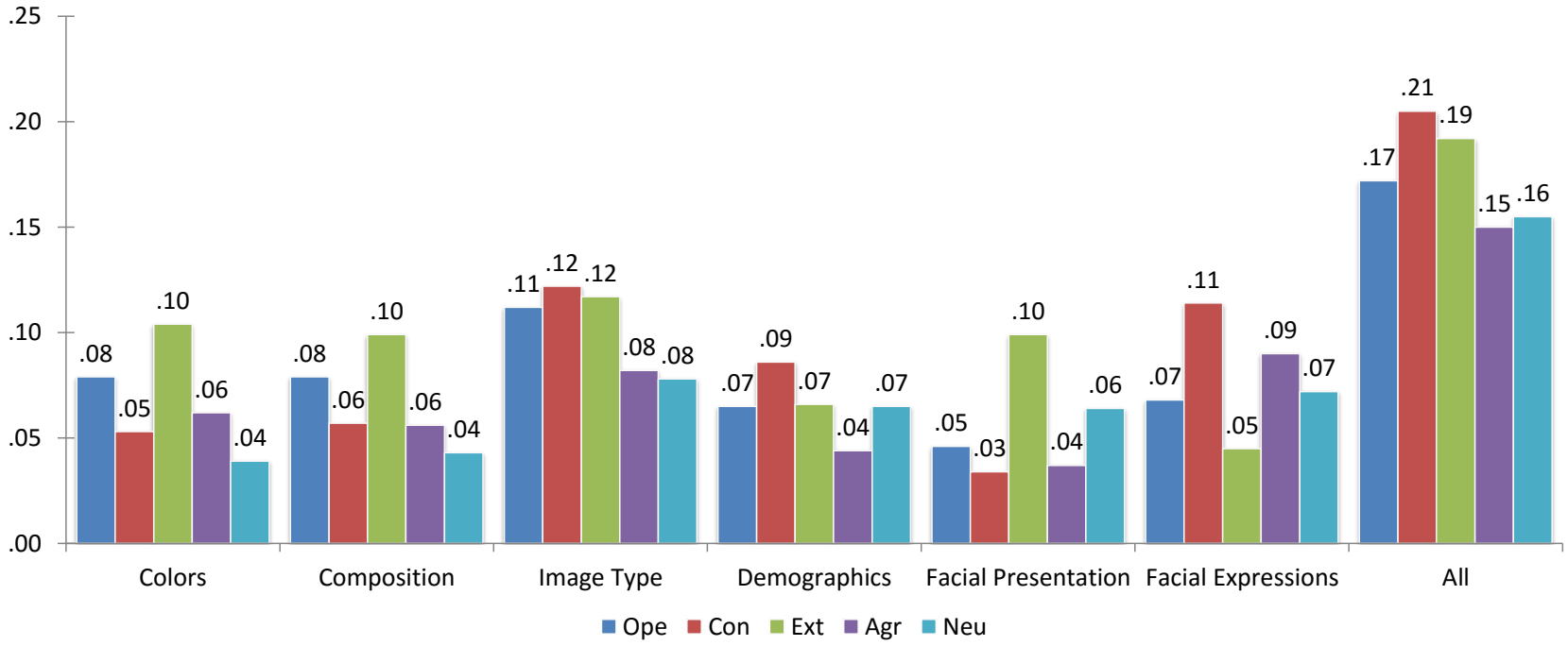
Overview - Agreableness

- photos are not artistic
- photos are not aesthetically pleasing
- most positive color emotions
- prefers faces
- expressing positive facial emotions

Overview - Extraversion

- photos are not artistic, but less than Agr
- photos are not aesthetically pleasing, but less than Agr
- positive color emotions
- prefers faces, especially multiple faces
- expressing positive facial emotions, less than Agr

Predictive Performance



TwitterText data set. Predictive performance using Linear Regression, measured in Pearson correlation over 10-fold cross-validation. All correlations are significant ($p < .05$, two-tailed t-test).

Take Aways

- Profile picture choice is influenced by personality
- Interpretable computer vision features hold significant prediction accuracy across all personality traits
- Text predicted personality is a good stand-in for survey personality that offers orders of magnitude statistical power

Challenges

We need to understand:

- stylistic text differences: phrase choice, readability, text complexity, syntax
- more psychological traits: empathy, all facets of personality, moral foundations and values
- role of biological factors: change in testosterone levels
- longitudinal change within individuals
- how to combine multiple modalities i.e. text, images, network
- difference between reality and perception of traits
- the goal of the personalisation e.g. trait match or complement

Thank you!

Resources

An analysis of the user occupational class through Twitter content.
Daniel Preoţiuc-Pietro, Vasileios Lampos, Nikolaos Aletras – ACL, 2015

Studying User Income through Language, Behaviour and Affect in Social Media.
Daniel Preoţiuc-Pietro, Svitlana Volkova, Vasileios Lampos, Yoram Bachrach, Nikolaos Aletras – PLoS ONE 10(9), 2015

Analyzing Personality through Social Media Profile Picture Choice.
Leqi Liu, Daniel Preoţiuc-Pietro, Zahra Riahi Samani, Mohsen Mohadam, Lyle Ungar – ICWSM, 2016

Exploring Stylistic Variation with Age and Income on Twitter.
Lucie Flekova, Lyle Ungar, Daniel Preoţiuc-Pietro – ACL, 2016

www.preotiuc.ro