Analyzing Connections Between User Attributes, Images, and Text

Laura Burdick*[1], Rada Mihalcea†[1], Ryan L. Boyd‡[2], and James W. Pennebaker§[3]

[1]Department of Computer Science and Engineering, University of Michigan (Bob and Betty Beyster Building, 2260 Hayward Street, Ann Arbor, MI 48109-2121)

[2]Department of Psychology, Lancaster University (Lancaster, United Kingdom, LA1 4YF)

[3]Department of Psychology, The University of Texas at Austin (SEA 4.208, 108 E. Dean Keeton Stop A8000, Austin, TX 78712-1043)

*wenlaura@umich.edu, (tel) 734-763-0503
†mihalcea@umich.edu
‡r.boyd@lancaster.ac.uk
§pennebaker@utexas.edu

Abstract

Background / Introduction. This work explores the relationship between a person's demographic/psychological traits (e.g., gender, personality) and self-identity images and captions.

Methods. We use a dataset of images and captions provided by $N \approx 1,350$ individuals, and we automatically extract features from both the images and captions.

Results. We identify several visual and textual properties that show reliable relationships with individual differences between participants. The automated techniques presented here allow us to draw interesting conclusions from our data that would be difficult to identify manually, and these techniques are extensible to other large datasets. Additionally, we consider the task of predicting gender and personality using both single-modality features and multimodal features.

Conclusions. We show that a multimodal predictive approach outperforms purely visual methods and purely textual methods. We believe that our work on the relationship between user characteristics and user data has relevance in online settings, where users upload billions of images each day (Meeker M, 2014. Internet trends 2014-Code conference. Retrieved May 28, 2014).

Keywords. Personality, Gender, Natural Language Processing, Computer Vision, Computational Social Science.

1 Introduction

Images have increasingly become a central part of most people's online ecosystem – people upload profile photos, create memes, and use images as a means of communication. In total, over 1.8 billion digital images are added to the internet each day [1]. This tremendous quantity of visual data has exciting potential to be used to gain a deeper understanding into the thoughts and behaviors of people. Since many of the images shared online are personalized by a user, studying them gives us insight into the user herself.

Specifically, in this work, we aim to present new, interpretable psychological insight into the ways that image attributes (such as objects, scenes, and faces) and language features (such as words and semantic categories) relate to personality and gender. We do this by using a dataset of $N \approx 1,350$ individuals, where each person has provided images and captions. From this dataset, we extract an extensive set of visual and textual features. We use these features to

identify relationships between these features and the individual traits of personality and gender. To show the strength of these relationships, we also briefly consider whether image and language features have predictive power for demographic/psychological characteristics. This work builds on the work presented in Wendlandt et al. [2]. Specifically, in this paper, we expand on our previous work by including a larger set of correlations between image and text features and personality traits, reporting results on a regression task not previously considered, validating our results on a second dataset collected at a later period of time, expanding our overview of previous related work, and expanding our analyses.

After examining related work, we begin by describing our dataset. We then explain the various methods used for analyzing images and text, showing significant correlations that are found. Finally, we use our visual and textual features to predict both personality and gender. We close with a discussion and conclusion.

1.1 Related Work

When studying individuals, we are often trying to get a general sense of who they are as a person. These types of evaluations fall under the broader umbrella of *individual differences*, a large area of research that tries to understand the various ways in which people are psychologically different from one another, yet relatively consistent over time [3]. A large amount of research in the past decade has been dedicated to the assessment and estimation of individual characteristics as a function of various behavioral traces. In our case, these traces are images and captions collected from undergraduate students.

The estimation of individual characteristics has been employed in various downstream tasks in fields such as public health [4] and politics [5, 6]. Some of the attributes targeted for extraction focus on demographic related information, such as gender/age [7, 8, 9, 10, 11, 12, 13, 14], race/ethnicity [15, 16, 17, 14], and location [18], yet other aspects are mined as well, among them emotion and sentiment [19], personality types [20, 21, 22, 23], user political affiliation and sentiment [24, 6, 25], mental health diagnosis [26], and even lifestyle choices such as coffee preference [15]. The task is typically approached from a machine learning perspective, with data originating from a variety of user-generated content, most often microblogs (from Twitter) [23, 4, 14], article comments to news stories or op-ed pieces [27], social posts (originating from sites such as Facebook, MySpace, Google+) [26], or discussion forums on particular topics [28]. Classification labels are then assigned either based on manual annotations (such as Amazon

Mechanical Turk [14]), self identified user attributes ("I am a 20 year old African American") [15], affiliation with a given discussion forum type, or online surveys set up to link a social media user identification to the responses provided (such as the embedded personality test survey application developed by Schwartz et al. [29]). Additional modeling information may surface from meta-data, such as geolocation provided by the Twitter API [16], or by applying distributions learnt from real world data, such as those collected as part of the US Census [30, 31], or by leveraging the social connections of a given user within a network [32, 33]. Learning has typically employed bag-of-words lexical features (n-grams) [12, 34, 35], with some works focusing on deriving additional signals from the underlying social network structure [15, 32, 33, 25], syntactic and stylistic features [36], or the intrinsic social media generation dynamic [25]. We should note that some works have also explored unsupervised approaches for demographic dimensions extraction, among them large-scale clustering [37] and probabilistic graphical models [38].

In our work, we focus on two individual attributes, namely personality and gender, and we highlight below the previous work on these tasks.

1.1.1 Personality

Much of the work in individual differences research focuses on the topic of *personality*. Generally speaking, "personality" refers to constellations of feelings, behaviors, and cognitions that co-occur within an individual and are relatively stable across time and contexts. Personality is most often conceived within the Big 5 personality framework, and these five dimensions of personality are predictive of important behavioral outcomes such as marital satisfaction [39] and even health [40].

From a computational perspective, the problem of identifying user personality has primarily been approached using Natural Language Processing (NLP) methods. While the textual component of our work focuses on short image captions, most previous research used longer bodies of text such as essays or social media updates [41]. N-grams, as well as psychologically-derived linguistic features such as those provided by LIWC, have been shown to have significant predictive power for personality [42, 43].

In addition to textual inference, there has been a recent movement towards incorporating images into the study of individual differences [44, 45].

1.1.2 Gender

Contemporary research on individual differences extends well beyond personality evaluations to include variables such as gender, age, life experiences, and so on – facets that differ between individuals but are not necessarily caused by internal psychological processes. In addition to personality, we also consider gender in this work.

As with personality, the computational inference of gender has primarily been approached using NLP techniques [8, 18, 46]. Relevant to the current work, however, is work by You et al. exploring the task of predicting gender given a user's selected images on Pinterest, an online social networking site [47]. In contrast to using data from a social network, this work uses data collected from students at a university. Additionally, we consider both gender and personality, while only gender is considered in You et al.

1.1.3 Inference from Multiple Modalities

Our work also relates to the recent body of research on the joint computational use of language and vision. Our multimodal predictive approach is particularly related to automatic image annotation, the task of extracting semantically meaningful keywords from images [48]. Other related multimodal approaches can be found in the fields of image captioning [49] and joint text-image embeddings [50]. Some of these approaches rely on very large visual and textual corpora. For example, Johnson et al. train an image captioning algorithm using Visual Genome, a dataset with greater than 94,000 images [51].

2 Methods

In this section, we describe the dataset that we use, as well as the visual and textual features that we extract.

2.1 Dataset

We use a dataset provided by James Pennebaker and Samuel Gosling at the University of Texas at Austin, collected from their Fall 2015 online undergraduate introductory psychology class.[1] The dataset includes free response data and responses to standard surveys collected from 1,353 students ages 16 to 46 (average 18.8 $\pm$ 2.10). The ethnicity distribution is

[1] This data was collected under IRB approval at UT Austin.

40.3% Anglo-Saxon/White, 27.1% Hispanic/Latino, 22.3% Asian/Asian American, 5.5% African American/Black, and 4.8% Other/Undefined.

Three elements of this dataset are of particular interest to our research:

2.1.1 Free Response Image Data

Each student was asked to submit and caption five images that expressed who he/she is as a person. The following prompt was used: *Please upload 5 different pictures that express who you are. They could be pictures of you, your friends, your possessions, or anything that you feel expresses your personality. Pick out the five pictures before you begin the assignment. Also, when you upload each picture, write a brief description or caption about it.*

As Fig 1 illustrates, students submitted a wide range of images, from memes to family photos to landscapes. Some students chose to submit fewer than five images. All images were converted to the JPG format and resized so that the longest edge of the image was 700 pixels; this preprocessing ensures efficient and uniform calculations across the dataset.

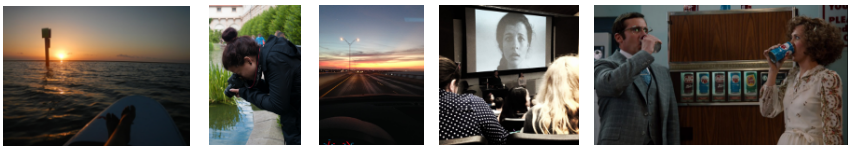

Figure 1: Five images from the dataset submitted by a single student (with student faces blurred out for privacy) [2]. The accompanying captions are: (A) I'd rather be on the water. (B) The littlest things are always so pretty (and harder to capture). (C) I crossed this bridge almost every day for 18 years and never got tired of it. (D) The real me is right behind you. (E) Gotta find something to do when I have nothing to say.

2.1.2 Big 5 Personality Ratings

The Big 5 personality dimensions include: *Openness* (example adjectives: artistic, curious, imaginative, insightful, original, wide interests); *Conscientiousness* (efficient, organized, planful, reliable, responsible, thorough); *Extraversion* (active, assertive, energetic, enthusiastic, outgoing, talkative); *Agreeableness* (appreciative, forgiving, generous, kind, sympathetic, trusting); and *Neuroticism* (anxious, self-pitying, tense, touchy, unstable, worrying) [52].

To measure these personality dimensions, each student completed the BFI-44 personality inventory, a self-report 44-question survey used to score individuals along each of the Big 5 personality dimensions using a 1-to-5 Likert scale [53]. Descriptive statistics for each personality

dimension within the current sample are presented in Table 1. Figure 2 shows correlations between the different dimensions of personality in our dataset. The highest positive correlation is between Agreeableness and Conscientiousness, while the largest negative correlation is between Extraversion and Neuroticism.

Table 1: Statistics for each personality dimension.

Dimension	Mean	Median	Std Dev
Openness	3.618	3.600	0.623
Conscientiousness	3.459	3.444	0.649
Extraversion	3.173	3.125	0.813
Agreeableness	3.716	3.778	0.644
Neuroticism	3.011	3.000	0.752

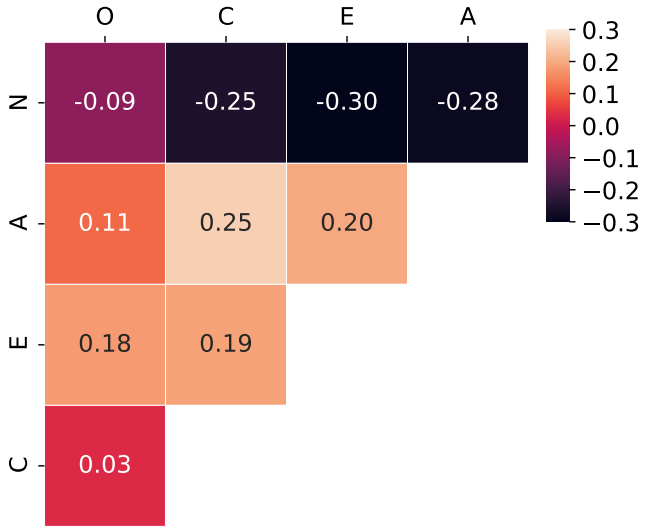

Figure 2: Pearson correlations between Big 5 personality dimensions in our dataset. O, C, E, A, and N stand for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

2.1.3 Gender

Finally, demographic data is also associated with each student, including gender, which we use in our work. The gender distribution in the dataset is 61.6% female, 37.8% male, and 0.5% undefined. Gender-unspecified students are omitted from our analyses.

2.1.4 Computing Correlations

An important contribution of our work is gathering new insights regarding textual and image attributes that correlate with personality and gender. Each of the personality dimensions is

continuous, therefore, a version of the Pearson correlation coefficient is used to calculate correlations between personality and visual and textual features. Since there are, in some cases, thousands of image or text features, we must account for inferential issues associated with multiple testing (e.g., inflated error rates); we address such issues using a multivariate permutation test [54].

This approach is done by first calculating the Pearson product-moment correlation coefficient r for two variables. Then, for a high number of iterations (in our case, 10,000), the two variables are randomly shuffled and the Pearson coefficient is re-calculated. At the end of the shuffling, a two-tailed p-test is conducted. Only when the original Pearson's r is found to be statistically significant in comparison to all of the random coefficients is the original result considered to be legitimate. As discussed in Yoder et al. [54], for small sample sizes, this multivariate permutation test has more statistical power than the common Bonferroni correction.

Unlike personality, gender is a categorical variable. Thus, Welch's t-tests are used to look for significant relationships between gender and image and text features. These relationships are measured using effect size (Cohen's d), which measures how many standard deviations the two groups differ by, and is calculated by dividing the mean difference by the pooled standard deviation.

2.2 Analyzing Images

In order to explore the relationship between images and psychological attributes, we want to extract meaningful and interpretable image features that have some connection to the user. In this section, we summarize both low-level raw visual features as well as high-level attributes such as the scene of an image, the number of faces in an image, and the objects in an image. How we extracted these features is described in more detail in previous work [2]. We use these features to explore significant correlations between image attributes and user attributes.

2.2.1 Raw Visual Features

In this section, we describe basic image statistics that can provide a good summary of the structural, color, and textual properties of an image, which in turn can provide insights into the attributes of the person submitting the image. Higher-level visual features are considered in following sections.

Colors. Past research has shown that colors are associated with abstract concepts [55]. For

instance, red is associated with excitement, yellow with cheerfulness, and blue with comfort, wealth, and trust. Furthermore, research has shown that men and women perceive color differently. In particular, one study found that men are more tolerant of gray, white, and black than are women [56].

To characterize the distribution of colors in an image, we classify each pixel as one of eleven named colors using the method presented by Van De Weijer et al. [57].

Brightness and Saturation. Images are often characterized in terms of their brightness and saturation. Here, we use the HSV color space, where brightness is defined as the relative lightness or darkness of a particular color, from black (no brightness) to light, vivid color (full brightness). Saturation captures the relationship between the hue of a color and its brightness and ranges from white (no saturation) to pure color (full saturation). We calculate the mean and the standard deviation for both the brightness and the saturation.

Previous work has also used brightness and saturation to calculate metrics measuring pleasure, arousal, and dominance [58].[2]

Texture. The texture of an image provides information about the patterns of colors or intensities in the image. Following Lovato et al. [60], we use Grey Level Co-occurrence Matrices (GLCMs) to calculate four texture metrics: contrast, correlation, energy, and homogeneity.

Static and Dynamic Lines. Previous work has shown that the orientation and width of a line can have various emotional effects on the viewer [59]. For example, diagonal lines are associated with movement and a lack of equilibrium. To capture some of these effects, we measure the percentage of static lines with respect to all of the lines in the image. Static lines are defined as lines that are within $\pi/12$ radians of being vertical or horizontal.

Circles. The presence of circles and other curves in images has been found to be associated with emotions such as anger and sadness [55]. Following the example of Redi et al. [55], we calculate the number of circles in an image.

Correlations. Once the entire set of raw features is extracted from the images, correlations between raw features and personality/demographic features are calculated. Table 2 presents significant correlations between visual features and personality traits. One correlation to note is a positive relationship between the number of circles in an image and extraversion. This is likely because the circle detection algorithm often counts faces as circles, and faces have a

[2]For prediction results, we use a slightly different version of dominance ($Dominance = 0.76y + 0.32s$), as formulated in Machajdik and Hanbury [59].

natural connection with the social facets of extraversion. Our results also validate the findings of Valdez and Mehrabian, who suggest that pleasure, arousal, and dominance have emotional connections [58]. Here we show that these metrics also have connections to personality.

Table 2: Significant correlations between raw visual features and Big 5 personality traits. These correlations are corrected using a multivariate permutation test, as described in the paper. O, C, E, A, and N stand for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

	Big 5 Personality Dimensions				
Image Attributes	O	C	E	A	N
Black	-	-	-	-	-0.06
Blue	-	-0.07	0.06	-	-
Grey	-	-	-0.11	0.06	-
Orange	-	-	0.07	-	-
Purple	-	-	0.06	-	-
Red	-	-	-	-0.06	-
Brightness Std Dev	-	-	0.07	-	-
Saturation Mean	-	-0.06	0.07	-	-
Saturation Std Dev	-	-0.06	0.06	-0.06	-
Pleasure	-	-0.05	0.07	-	-
Arousal	-	0.06	-	-	-
Dominance	-	-	-0.06	-0.02	0.06
Homogeneity	0.05	-	-	-	-
Static Lines %	-	-	-0.07	-	-
Num of Circles	-	-0.06	0.10	-	-

Table 3 shows effect sizes for features significantly different between men and women. As suggested by previous research, men are more likely to use the color black [56]; other correlations appear to confirm stereotypes, e.g., a stronger preference by women for pink and purple.

Table 3: Raw visual features where there is a significant difference ($p < 0.05$) between male and female images. Positive effect sizes indicate that women prefer the feature, while negative effect sizes indicate that men prefer the feature.

Image Attributes	Effect Size
Pink	0.455
Static Lines %	-0.360
Black	-0.325
Brightness Mean	0.266
Saturation Std Dev	-0.176
Purple	0.167
Brown	0.166
Homogeneity	0.118
Red	0.111

2.2.2 Scenes

Previous research has linked personal spaces (such as bedrooms and offices) with various personality attributes, indicating that how people compose their spaces provides clues about their psychology, particularly through self-presentation and related social processes [61].

In order to identify the scene of an image, we use Places-CNN [62], a convolutional neural network (CNN) trained on approximately 2.5 million images and able to classify an image into 205 scene categories. To illustrate, Fig 3 shows two classified images. For each image, we use the softmax probability distribution over all scenes as a feature vector.

Figure 3: Top scene classifications for two images, along with their probabilities [2]. A: Coffee Shop (0.53), Ice Cream Parlor (0.24). B: Parking Lot (0.57), Sky (0.26).

Correlations. Scenes strongly correlated with personality traits are shown in Table 4. The strongest positive correlation is between extraversion and ballrooms, and the strongest negative correlation is between extraversion and home offices. Findings such as these are conceptually sound, as individuals tend to engage in personality-congruent behaviors. In other words, individuals scoring high on extraversion are expected to feel that inherently social locations, such as ballrooms, are more relevant to the self than locations indicative of social isolation, such as home offices.

We also measure the relationship between scenes and gender. Table 5 shows scenes that are associated with either males or females. Men are more commonly characterized by sports-related scenes, such as football and baseball stadiums, whereas women are more likely to have photos from ice cream and beauty parlors. As illustrated in Fig 3, the scene detection algorithm tends to conflate coffee shops and ice cream parlors, so this observed preference for ice cream parlors could be partially attributed to a preference for coffee shops.

Table 4: Significant correlations between scene features and Big 5 personality traits. Only correlations with $p \geq 0.07$ are shown. These correlations are corrected using a multivariate permutation test, as described in the paper. O, C, E, A, and N stand for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

	Big 5 Personality Dimensions				
Image Attributes	O	C	E	A	N
Art Studio	-	-0.09	-	-	-
Auditorium	-	-	0.08	-	-
Ballroom	-	-	0.12	-	-
Baseball Field	-0.08	-	-	-	-
Beauty Salon	-	-	-	-	0.08
Bedroom	-	-	-0.08	-	-
Boardwalk	-	-	-	0.08	-
Bookstore	-	0.08	-0.11	-	-
Botanical Garden	-	0.08	-	-	-
Bus Interior	-	-	-	-0.08	-
Butte	0.08	-	-	-	-
Canyon	-	-	-	0.11	-
Coffee Shop	-	-0.08	-	-	-
Creek	0.08	-	-	-	-
Fire Station	-0.08	-	-	-	0.08
Formal Garden	-	0.09	-	-	-
Fountain	-	0.08	-	-	-
Game Room	-	-0.08	-	-	-
Home Office	-	-	-0.12	-	-
Hot Spring	0.08	-	-	-	-
Mansion	-	-	-	0.10	-
Martial Arts Gym	-0.09	-	-	-	-
Museum Indoor	-	-	-	-0.08	-
Pantry	-	-	-0.11	-0.10	-
Pavilion	-	-	-	-	-
Phone Booth	-0.08	-	-	-	-
Playground	-	-	-	0.09	-
Restaurant	-0.09	-	-	-	-
River	-	0.09	-	-	-
Shower	-	-	-0.09	-	-
Stadium Baseball	-0.08	-	-	-	-
Valley	0.08	-	-	-	-
Veranda	-	-	-	0.08	-

Table 5: Scene features where there is a significant difference ($p < 0.05$) between male and female images. Only features with an effect size of magnitude > 0.07 are shown. Positive effect sizes indicate that women prefer the feature, while negative effect sizes indicate that men prefer the feature.

Image Attributes	Effect Size
Beauty Salon	0.168
Ice Cream Parlor	0.156
Office	-0.133
Slum	0.131
Football Stadium	-0.130
Basement	-0.109
Art Studio	0.105
Herb Garden	0.103
Music Studio	-0.102
Baseball Stadium	-0.102
Gas Station	-0.101
Game Room	-0.099
Vegetable Garden	0.097
Botanical Garden	0.096
Yard	0.096
Conference Room	-0.094
Engine Room	-0.094
Home Office	-0.092
Reception	-0.091
Assembly Line	-0.090
Bedroom	0.088
Television Studio	-0.083
Baseball Field	-0.080
Office Building	-0.080
Butcher's Shop	0.079
Playground	0.075
Hot Spring	0.074
Nursery	0.073
Shoe Shop	-0.073

2.2.3 Faces

Most aspects of a person's personality are expressed through their social behaviors, and the number of faces in an image can capture some of this behavior. We use the work by Mathias et al. to detect faces in images [63].

Correlations. There is a strong positive correlation ($r = 0.17$) between the number of faces and extraversion, which is intuitive because extraverts are often thought of as enjoying social activities. There are also positive correlations between faces and openness ($r = 0.08$) and neuroticism ($r = 0.11$), while there is a negative correlation between faces and agreeableness ($r = -0.07$). With respect to gender, women have significantly more faces in their images than men (effect size = 0.160).

2.2.4 Objects

Previous research has indicated that people can successfully predict other people's personality traits by observing their possessions [64]. This indicates that object detection has the ability to capture certain psychological insight. Fig 4 shows an example image with several detected objects.

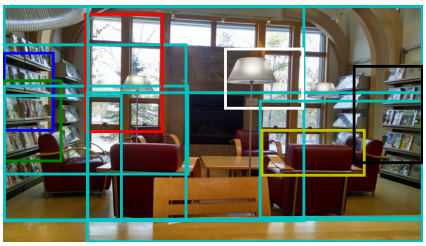

Figure 4: Several detected objects and their bounding boxes in an image of a library [2]. Pictured objects are libraries (cyan), plate racks (blue), tobacco shops (green), window screens (red), table lamps (white), dining tables (yellow), and bookcases (black).

Because of the small size of our dataset and the large number of ImageNet objects, this feature vector is somewhat sparse and hard to interpret. To increase interpretability, we consider two coarser-grained systems of classification: WordNet supersenses and WordNet domains. WordNet [65] is a large hierarchical database of English concepts (or synsets), and each ImageNet object is directly associated with a WordNet concept. Supersenses are broad semantic classes labeled by lexicographers [66], and eight supersenses are present in our set of ImageNet objects: communication, object, plant, food, artifact, animal, substance, and person. WordNet domains [67] is a complementary synset labeling. It groups WordNet synsets into various

domains, such as medicine, astronomy, and history. The domain structure is hierarchical, but here we consider only basic WordNet domains, which are domains that are broad enough to be easily interpretable. An object is allowed to fall into more than one domain. Table 6 lists both the WordNet supersenses and the WordNet domains and provides a few examples of each category.

Correlations. WordNet supersenses and WordNet domains correlate significantly with multiple personality traits, as shown in Table 7. We begin to see some patterns emerge across different domains. For example, multiple technical disciplines (engineering, telecommunication, physics) are negatively correlated with both conscientiousness and agreeableness. There are very few correlations with neuroticism, which is something that we observe with other image features as well.

Table 8 shows object classes that are different for males and females. These object classes connect back to scenes associated with men and women. For example, men are more likely to have sports objects in their images, reflected in the fact that men are more likely to include scenes of offices and sports stadiums.

2.3 Captions

When available, captions can be considered another way of representing image content via a textual description of the salient objects, people, or scenes in the image. Importantly, the captions have been contributed by the same people who contributed the images, and therefore they represent the views that the image "owners" have about their content. How we extracted these features is described in more detail in previous work [2].

2.3.1 Stylistic Features

To capture writing style, we consider surface-level stylistic features, such as the number of words and the number of words longer than six characters. We also use the Stanford Named Entity Recognition system to extract the number of references to people, locations, and organizations [68].

Finally, we look at readability and specificity metrics. Readability scores are usually based on the length and difficulty of words and sentences, and they capture how hard the text is to comprehend. We consider a variety of metrics: Flesch Reading Ease (FRE), Automated Readability Index (ARI), Flesch-Kincaid Grade Level (FK), Coleman-Liau Index (CLI), Gun-

Table 6: Selected examples for each WordNet supersense and basic WordNet domain.

WordNet Supersenses	
Communication	Web site, comic book, traffic light
Object	Alp, bubble, cliff
Plant	Rapeseed, daisy, corn
Food	Menu, plate, guacamole, trifle
Artifact	Abacus, bakery, breastplate
Animal	Mud turtle, airedale, meerkat
Substance	Toilet tissue
Person	Ballplayer, groom, scuba diver
Basic WordNet Domains	
History	Breastplate, cuirass, pickelhaube
Art	Paintbrush, violin, Polaroid camera
Religion	Church, monastery, mosque
Radio and TV	Radio, screen, television
Play	Golf ball, jigsaw puzzle, punching bag
Sport	Baseball, basketball, golf cart
Agriculture	Harvester, plow, thresher
Food	Menu, eggnog, acorn
Home	Bath towel, broom, dishwasher
Architecture	Triumphal arch, library, traffic light
Computer Science	Computer keyboard, mouse, web site
Engineering	Pier, oscilloscope, remote control
Telecommunication	Loudspeaker, cellular telephone, pay-phone
Medicine	Medicine chest, neck brace, Band Aid
Astronomy	Radio telescope
Biology	Tench, goldfish, daisy
Animals	Walker hound, sea cucumber, wood rabbit
Chemistry	Face powder, French loaf, cauliflower
Plants	Granny Smith, strawberry, coral fungus
Earth	Alp, cliff, coral reef
Mathematics	Abacus
Physics	Gasmask, whistle, oscilloscope
Anthropology	Maypole
Health	Face powder, hair spray, lipstick
Military	Assault rifle, bow, breastplate
Publishing	Ballpoint, binder, fountain pen
Artisanship	Hammer, plane, thimble
Commerce	Grocery store, hand blower, restaurant
Industry	Carpenter's kit, chain, lumbermill, power drill
Transport	Ambulance, missile, seat belt
Economy	Slot, streetcar, lumbermill
Administration	File
Law	Guillotine, prison
Tourism	Cab, triumphal arch, volcano
Fashion	Apron, bow tie, cowboy boot

Table 7: Significant correlations between object features and Big 5 personality traits. These correlations are corrected using a multivariate permutation test, as described in the paper. O, C, E, A, and N stand for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

	Big 5 Personality Dimensions				
Image Attributes	O	C	E	A	N
WordNet Supersenses					
Animal	-	-	0.063	-	-
Person	-	-	-	-	-0.58
Basic WordNet Domains					
History	-	-	0.06	-	-
Play	-0.10	-	-	-	-
Sport	-0.10	-	-	-	-
Home	-	-0.06	-0.09	-	-
Engineering	-	-0.06	-0.07	-0.08	-
Telecommunication	-	-0.07	-	-0.08	-
Astronomy	-	-	-	-0.06	-
Biology	-	-	0.07	-	-
Animals	-0.08	-	-	-	-
Physics	-	-0.08	-	-0.09	-
Anthropology	-	0.06	-	-	-
Artisanship	-	-0.06	-	-	-
Industry	-	-	-0.08	-	-
Transport	-	-	-	-	-0.07
Economy	-	-	-0.08	-	-
Fashion	-0.07	0.06	0.11	0.05	-

Table 8: Object features where there is a significant difference ($p < 0.05$) between male and female images. Positive effect sizes indicate that women prefer the feature, while negative effect sizes indicate that men prefer the feature.

Image Attributes	Effect Size
WordNet Supersenses	
Artifact	-0.213
Person	-0.173
Food	0.107
Basic WordNet Domains	
Sport	-0.235
Play	-0.231
Transport	-0.186
Military	-0.172
Animals	-0.155
History	-0.153
Art	-0.142
Chemistry	0.140
Food	0.136
Plants	0.121

ning Fog Index (GFI), and SMOG score (SMOG). Specificity refers to how much detail a text contains. We calculate this using the Speciteller system [69].

2.3.2 N-grams

In addition to style, we want to capture the content of each caption. We do this by considering unigrams, bigrams, and trigrams.

2.3.3 LIWC Features

Linguistic Inquiry and Word Count (LIWC) is a word-based text analysis program [42]. It focuses on emotional, cognitive, and social processes, as well as broad categories such as language composition. We analyze each piece of text using LIWC in order to capture psychological dimensions of writing. For each of the 86 LIWC categories, we calculate a feature that reflects the percentage of words belonging to that category which are present in the caption.

2.3.4 MRC Features

The MRC Psycholinguistic Database contains statistics about word use [70]. MRC features are calculated by averaging the values of all of the words in a caption. Specifically, in our analysis, certain MRC features emerge as particularly relevant. These include several features suggested by Kucera and Francis, including word frequency counts, which capture how common a word is in standard English usage. We also see measures for meaningfulness, imagery, and length (e.g., number of letters, phenomes, and syllables). These features provide a complementary perspective to the LIWC features.

2.3.5 Word Embeddings

For prediction purposes, we also consider each word's embedding. Word2vec (w2v) is a method for creating a multidimensional embedding for a particular word [71]. Google provides pre-trained word embeddings on approximately 100 billion words of the Google News dataset.[3] For each caption in our dataset, we average together all of the word embeddings to produce a single feature vector of length 300. We use the Google embeddings for this, discarding words that are not present in the pre-trained embeddings.

[3]Available at https://code.google.com/archive/p/word2vec/.

2.3.6 Correlations

For our analysis, all text features are normalized by word count. Table 9 shows correlations between language features and personality. Interestingly, there are very few strong correlations for extraversion. This is complementary to what we see with images, where there are many strong correlations for extraversion, suggesting that we are gleaning different aspects of personality from both images and text.

Table 10 shows language features that are different between men and women. Things to note here are that women tend to write longer captions and men again exhibit a preference for talking about sports.

3 Results

In this section, we consider single-modality and multimodal prediction tasks.

3.1 Multimodal Prediction

The task of prediction can provide valuable insights into the relationship between images, captions, and demographic or psychological dimensions. In this section, we consider both single modality features and multimodal features. We also address two prediction tasks: classification and regression. Classification is a more coarse-grained task, and is the typical prediction task considered in previous work for such latent user dimensions. On the other hand, regression is more fine-grained and produces more nuanced results.

3.1.1 Single Modality Methods

To understand the predictive power of images and captions individually, we consider a series of predictions using feature sets derived from either only visual data or only textual data. These feature sets are the same features that we described above.

To gain insight into whether textual and visual data complement each other, Fig 5 shows correlations between attributes predicted using only image features and attributes predicted using only text features. The low correlations between visual and textual predictions of the same trait indicate that images and text are capturing different aspects of each trait and have the potential to be used together to gain a fuller picture. We explore the joint use of images and text below.

Table 9: Significant correlations between language attributes and Big 5 personality traits [2]. All features except for the word count itself are normalized by the word count. Only unigrams, LIWC categories, and MRC categories that have one of the top five highest correlations or one of the top five lowest correlations are shown. These correlations are corrected using a multivariate permutation test, as described in the paper. O, C, E, A, and N stand for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

	Big 5 Personality Dimensions				
Language Attributes	O	C	E	A	N
Stylistic Features					
Number of words	-	0.14	-	-	0.07
Words longer than six chars	-	0.09	0.06	-	-
Number of locations	-	0.07	-	-	-0.07
Readability - FRE	-0.13	-	-	-	-
Readability - ARI	-	0.06	-	-	-
Readability - GFI	-0.14	-	-	-	-
Readability - SMOG	-0.13	-	-	-	-0.06
Specificity	-	0.08	-	-	-0.06
Unigrams					
Decid	-	-0.12	-	-	-
Diff	0.11	-	-	-	-
In	0.11	-	-	-	-
It	0.11	-	-	-	-
King	-	0.06	-	-0.15	-
Level	-	-	-0.12	-	-
My	-0.14	0.07	-	-	-
Photoshop	0.10	-	-	-	-
Sport	-0.14	-	-	-	-
Writ	0.10	-	-	-	-
LIWC Categories					
Achievement	-	0.08	-	-	-
All Punctuation	-	0.08	-	-	-0.07
Discrepancies	-	-	0.10	-	-0.07
1st person singular personal pronouns	-0.10	-	-	-	-
Inclusive	-	-	-	0.08	-
Occupation	-	0.08	0.06	-	-
Other References	-0.10	-	-	-	-0.06
1st person personal pronouns	-0.10	-	-	-	-
Sports	-0.11	0.07	-	-	-
Unique	-	0.07	-	0.08	-0.09
MRC Categories					
Imagery	-0.07	0.06	-	0.06	-0.07
Kucera-Francis Num of Categories	-0.07	0.06	-	0.07	-0.09
Kucera-Francis Num of Samples	-0.08	-	-	-	-0.07
Mean Pavio Meaningfulness	-0.08	-	-	-	-0.07
Num of Letters in Word	-	0.08	-	0.07	-0.08
Num of Phonemes in Word	-	0.08	-	0.07	-0.08
Num of Syllables in Word	-	0.08	-	0.08	-0.08

Table 10: Language features where there is a significant difference ($p < 0.05$) between male and female images [2]. All features except for the word count itself are normalized by the word count. Only unigrams that have one of the top ten effect sizes (by magnitude) are shown. Positive effect sizes indicate that women prefer the feature, while negative effect sizes indicate that men prefer the feature.

Feature	Effect Size
Stylistic Features	
Number of Words	0.174
Readability - GFI	-0.161
Readability - SMOG	0.146
Readability - FRE	-0.136
Unigrams	
Boyfriend	0.361
Girlfriend	-0.360
Was	0.287
Play	-0.285
She	0.264
Them	0.262
Sport	-0.254
Sist	0.244
Gam	-0.242
Enjoy	-0.236
LIWC Categories	
Prepositions	-0.200
Past Focus	0.176
Sports	-0.173
Work	-0.167
Period	-0.157
Other References	0.145
Quote	-0.133
Other	0.123
1st person plural personal pronouns	0.123
MRC Categories	
Kucera-Francis Written Freq.	-0.139
Kucera-Francis Num of Samples	-0.134

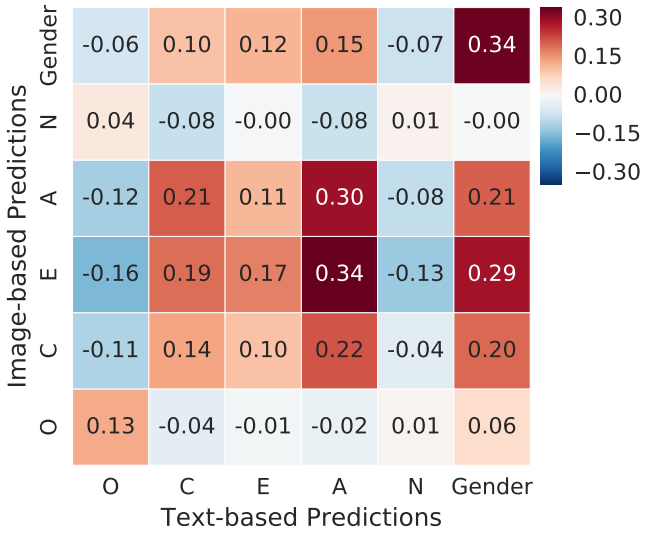

Figure 5: Pearson correlations between traits predicted using only image attributes and traits predicted using only text attributes [2]. Big 5 personality traits are denoted by O, C, E, A, and N. These predictions are done using 1,346 people (75% training, 25% test) and a random forest regressor with 500 trees.

3.1.2 Multimodal Methods

We experiment with several methods of combining visual and textual data. First, we concatenate both the image and text feature vectors (excluding w2v embeddings). In the following results tables, this is labeled as *All* row in the *Image and Caption Attributes* section.

To provide a more nuanced combination of features, we introduce the idea of image-enhanced unigrams (IEUs). This is a bag-of-words representation of both an image and its corresponding caption. It includes all of the caption unigrams as well as unigrams derived from each image. We consider two methods, macro and micro, for generating image unigrams. For the macro method, we examine each individual image. If a color covers more than one-third of the image, the name of the color is added to the bag-of-words. The scene with the highest probability and any objects detected in the image are also added. The unigrams from each individual image are then combined with the caption unigrams to form the set of macro IEUs. To generate micro IEUs, we reverse the process. First, we aggregate the image feature vectors into a single vector, and then we extract the image unigrams and combine them with the caption unigrams.

We use IEUs in several different ways for prediction. First, we consider them both in isolation and concatenated with all of the previous visual and textual features (excluding w2v). We also explore using the pre-trained w2v model to represent the IEUs and produce richer embeddings. Instead of only averaging together the embeddings of each caption unigram, we average together

the embeddings of each IEU. Finally, we consider these enriched embeddings concatenated with all of the previous visual and textual features.

A significant advantage of these multimodal approaches is that they can be used with relatively small corpora of images and text. Large background corpora are used for training (e.g., for training the scene CNN), but these models have already been trained and released. Our approaches work when there is only a small amount of training data, as is often the case when ground truth labels are expensive to obtain (e.g., when these labels come from a survey, as in our case). This is demonstrated on our dataset, which consists of short captions and a relatively limited set of images.

3.1.3 Classification Results

In order to assess the different prediction methods, we consider six coarse-grained classification tasks, one for each personality trait and one for gender. For each prediction, we divide the data into high and low segments. The high segment includes any person who has a score greater than half a standard deviation above the mean, while the low segment includes any person who has a score lower than half a standard deviation below the mean. All other data points are discarded. In doing these coarse-grained classification tasks, we follow previous work [72, 43], which suggested that classification serves as a useful approximation to continuous rating.

We use a random forest with 500 trees and 10-fold cross validation across individuals in the dataset. Table 11 shows the classification results. As a baseline, we include a model that always predicts the most common training class.

Results using individual modalities are shown in the top part of Table 11. The prediction results show that image features in isolation are able to significantly classify both extraversion and gender. Text features are also able to significantly classify these traits, with slightly less accuracy than image features. Text features have additional predictive power for openness.

The results obtained with the multimodal methods are shown in the bottom part of Table 11. As seen in the table, the methods using IEUs achieve the best results and are able to significantly classify both neuroticism and agreeableness, something that neither visual features nor textual features are able to do in isolation.

The features that we are able to classify with the highest accuracy are openness and agreeableness. This could be related to the fact that these two traits are positively correlated in our dataset (Figure 2).

Table 11: Classification accuracy percentages. * indicates significance with respect to the baseline ($p < 0.05$). O, C, E, A, and N stand for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

	Predicted Attributes					
Feature Set Used	O	C	E	A	N	Gender
Baseline: Most Common Class	51.4	52.0	49.2	51.8	52.3	59.8
Image Attributes Only						
Raw	49.7	47.3	53.9	51.6	52.2	61.5
Color	49.8	51.7	52.1	51.8	53.6	62.2
Object	**55.6**	51.7	57.2*	51.3	51.7	64.7*
Scene	55.5	53.8	59.8*	55.0	55.2	66.8*
Face	50.7	51.2	58.5*	54.1	51.1	59.7
All	54.8	**54.3**	**59.9***	**55.3**	**55.3**	**68.6***
Caption Attributes Only						
Stylistic	60.5*	48.0	50.4	50.8	51.0	58.6
Unigrams	60.2*	53.1	54.3	54.2	53.4	**67.6***
Bigrams	58.0*	**53.2**	**57.6***	53.4	**57.3**	65.1*
Trigrams	56.2*	51.0	55.5*	50.5	54.9	61.7
POS Unigrams	58.0*	49.3	50.0	52.9	56.1	60.2
POS Bigrams	57.5*	48.9	50.9	50.7	53.9	61.0
POS Trigrams	55.7	50.0	52.2	48.4	56.7	61.0
LIWC	59.6*	**53.2**	54.1	53.4	54.2	64.9*
MRC	55.4	49.4	50.9	52.4	52.4	60.8
All (except pre-trained w2v)	61.2*	52.2	53.3	54.6	55.2	65.1*
Pre-trained w2v (caption only)	**61.8***	51.4	55.4	**55.4**	56.5	67.1*
All + Pre-trained w2v (caption only)	61.2*	52.3	55.5	53.0	56.1	65.6*
Image and Caption Attributes						
All	60.5*	55.1	57.9*	55.3	56.8	67.1*
Macro IEU	58.5*	56.6	58.5*	54.2	54.7	**71.0***
Micro IEU	58.7*	54.4	58.9*	54.0	52.7	**71.0***
All + Macro IEU	60.0*	**57.1**	58.3*	54.2	56.9	68.1*
All + Micro IEU	59.1*	55.6	60.3*	54.8	**58.3***	69.1*
Pre-trained w2v (w/ Micro IEU)	**61.4***	54.8	59.6*	56.4*	56.5	68.6*
Pre-trained w2v (w/ Macro IEU)	61.0*	55.6	**60.5***	**57.0***	56.6	69.0*
All + Pre-trained w2v (w/ Micro IEU)	59.5*	54.8	59.1*	55.3	55.3	70.1*
All + Pre-trained w2v (w/ Macro IEU)	**61.4***	54.7	59.4*	55.2	56.5	70.8*

To enable direct comparison to previously published results, we also use our data to re-train the models used by Mairesse et al. to predict personality [43]; the re-trained classifier with the highest accuracies on our data, SMO, is shown in Table 12. We also include the relative error rate reduction between this model and our highest multimodal result. As shown in Table 12, our best multimodal approach outperforms the method from Mairesse et al., achieving relative error rate reductions between 5% and 16% across all categories. It is true that the relative error rate reduction for openness and neuroticism is smaller than the other dimensions, but in general we see that we are able to improve over Mairesse et al.

Table 12: Comparison between our best classification model and the best model (SMO) from Mairesse et al. [2]. * indicates significance with respect to the baseline ($p < 0.05$). The relative error rate reduction is between our model and the model from Mairesse et al. O, C, E, A, and N stand for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

	Predicted Attributes					
Feature Set Used	O	C	E	A	N	Gender
Baseline: Most Common Class	51.4	52.0	49.2	51.8	52.3	59.8
Mairesse et al.: SMO	59.1*	51.3	53.3	54.4	54.7	63.0
Our model: All + Pre-trained w2v (w/ Macro IEU)	61.0*	55.6	60.5*	57.0*	56.6	69.0*
Relative error rate reduction	4.6%	8.8%	15.4%	5.7%	4.2%	16.2%

3.1.4 Regression Results

Regression is a more fine-grained, and therefore more difficult, task than classification. We consider it because it gives us a more nuanced view of the effectiveness of our methods.

We use a random forest regressor with 500 trees and 10-fold cross-validation across individuals. As a baseline, we include a model that always predicts the average value of the training data. Table 13 reports r^2 scores. We notice patterns similar to the ones observed in the classification results. As before, image features alone are able to significantly predict both extraversion and gender, while text features are able to significantly predict openness, extraversion, and gender. Again, multimodal approaches outperform purely textual and purely visual approaches. Here, multimodal methods are able to significantly predict five out of the six categories, failing only to predict conscientiousness.

As we did for classification, we use our data to re-train the regression models used in Mairesse et al. [43]. Results for the regressor with the highest scores on our data, REPTree, is shown in Table 14, along with the relative error rate reduction between this model and our highest

Table 13: Regression r^2 scores. * indicates significance with respect to the baseline ($p < 0.05$). O, C, E, A, and N stand for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

	Predicted Attributes					
Feature Set Used	O	C	E	A	N	Gender
Baseline: Average Value	-0.011	-0.011	-0.007	-0.06	-0.004	-0.004
Image Attributes Only						
Raw	-0.044	-0.046	-0.026	-0.035	-0.012	0.036*
Color	-0.038	-0.048	-0.017	-0.044	-0.021	0.043*
Object	-0.057	-0.101	-0.030	-0.076	-0.097	0.043
Scene	-0.005	**-0.002**	0.021	-0.013	-0.007	0.121*
Face	-0.039	-0.036	-0.014	-0.019	-0.027	-0.040
All	**0.009***	-0.006	**0.033***	**-0.009**	**-0.000**	**0.145***
Caption Attributes Only						
Stylistic	-0.020	-0.058	-0.077	-0.049	-0.066	-0.016
Unigrams	0.050*	**0.012**	-0.001	-0.011	-0.010	0.169*
Bigrams	0.004	-0.024	0.007	-0.044	-0.037	0.104*
Trigrams	-0.049	-0.078	-0.053	-0.082	-0.109	0.021
POS Unigrams	-0.005	-0.020	-0.024	-0.021	-0.025	0.016
POS Bigrams	0.007	-0.055	-0.015	-0.019	-0.018	0.033*
POS Trigrams	-0.008	-0.052	-0.014	-0.022	-0.014	0.023
LIWC	0.057*	-0.005	0.005	-0.013	-0.010	0.101*
MRC	0.001	-0.047	-0.047	-0.015	-0.042	0.014
All (except pre-trained w2v)	0.055*	-0.003	**0.021***	-0.003	-0.010	0.164*
Pre-trained w2v (caption only)	0.065*	0.006	0.018	**-0.001**	**0.003**	0.121*
All + Pre-trained w2v (caption only)	**0.079***	0.011	**0.021***	-0.003	-0.000	**0.178***
Image and Caption Attributes						
All	0.048*	0.003	0.038*	-0.002	-0.001	0.203*
Macro IEU	0.028*	0.010	0.027	-0.013	-0.014	0.204*
Micro IEU	0.025	-0.002	0.033	-0.030	-0.035	0.197*
All + Macro IEU	0.057*	0.004	0.043*	-0.003	-0.001	0.208*
All + Micro IEU	0.054*	0.004	0.034*	-0.001	-0.003	0.207*
Pre-trained w2v (w/ Macro IEU)	0.045*	0.008	0.053*	**0.016***	0.004	0.176*
Pre-trained w2v (w/ Micro IEU)	0.048*	0.011	0.050*	0.017	-0.000	0.159*
All + Pre-trained w2v (w/ Macro IEU)	**0.063***	0.015	**0.057***	0.005	**0.005**	**0.224***
All + Pre-trained w2v (w/ Micro IEU)	0.060*	**0.016**	0.056*	0.010	-0.003	0.222*

multimodal result. When calculating error rates, r^2 scores are bounded below at zero, which is the expected performance of a baseline algorithm. Again, we see an improvement over Mairesse et al. Our best method achieves error rate reductions between 0.5% and 20%.

Table 14: Comparison between our best regression model and the best model (REPTree) from Mairesse et al. * indicates significance with respect to the baseline ($p < 0.05$). The relative error rate reduction is between our model and the model from Mairesse et al. O, C, E, A, and N stand for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

	Predicted Attributes					
Feature Set Used	O	C	E	A	N	Gender
Baseline: Average Value	-0.011	-0.011	-0.007	-0.06	-0.004	-0.004
Mairesse et al.: REPTree	0.012	-0.049	-0.018	-0.013	-0.015	0.027*
Our model: All + Pre-trained w2v (w/ Macro IEU)	0.063*	0.015	0.057*	0.005	0.005	0.224*
Relative error rate reduction	5.2%	1.5%	5.7%	0.5%	0.5%	20.2%

4 Discussion

In order to explore the replicability of these results, we gather another dataset, again from an online undergraduate introductory psychology class at the University of Texas at Austin. This second dataset was collected in Winter 2016, and contains 711 students. A comparison of the features extracted for the 2015 and 2016 data shows that for most of the features, the means and standard deviations are comparable.

We train a classifier on 2015 data and test it on 2016 data to evaluate how general the features that we extract are. To build the classifier, for each personality trait, students are split into a high group (where that trait is greater than a standard deviation above the mean) and a low group (where that trait is less than a standard deviation below the mean), discarding everyone who falls in the middle. The high group and the low group are balanced using under-sampling for each trait, and a random forest classifier with 500 trees is trained using ten-fold validation. The classifier is trained on all of the features from the 2015 data (excluding n-grams and part-of-speech n-grams) and tested on the 2016 data. Table 15 shows the results. In general, the classifier is able to beat the baseline, though not always significantly. However, this indicates that the features being used carry some information about the underlying personality and demographic traits.

The small size of both the 2015 and 2016 datasets could explain the lack of statistical significance in these results. It is possible that because both datasets are small, there is topic

Table 15: Precision-recall AUC for a classifier trained on 2015 data and tested on 2016 data. The standard deviation over ten folds is shown, and significant increases over the baseline are in bold. O, C, E, A, and N stand for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, respectively.

	Predicted Attributes					
	O	C	E	A	N	Gender
Baseline	0.50 ± 0.07	0.50 ± 0.11	0.50 ± 0.07	0.50 ± 0.10	0.50 ± 0.06	0.40 ± 0.03
Random Forest	0.58 ± 0.08	0.56 ± 0.09	0.56 ± 0.10	0.51 ± 0.12	0.57 ± 0.07	$\mathbf{0.52 \pm 0.04}$

shift between the two datasets, causing some of the 2015 results to be irrelevant on the 2016 dataset.

5 Conclusion

This research, using a new dataset of captioned images associated with user attributes, we have extracted a large set of visual and textual features and identified significant correlations between these features and the user traits of personality and gender. The automated techniques used to derive these features and find significant relationships are broadly applicable to other large visual and textual datasets. Specifically, in the domain of online communities, massive amounts of data are available. Some of these communities, like Pinterest, rely exclusively on visual content, while other communities, like Facebook and Twitter, include more textual content. We show how to automatically analyze this data and find meaningful psychological relationships. These techniques are not limited to the user dimensions of personality and gender and could be extended to other dimensions, such as age, education level, or location.

We have demonstrated the effectiveness of these image features in predicting user attributes; we believe this result can have applications in many areas of the web where textual data is limited. Finally, we have shown that a multimodal predictive approach outperforms purely visual methods and purely textual methods. Our multimodal methods are also effective on a relatively small corpus of images and text, which is useful in situations where data is limited.

6 Acknowledgements

This material is based in part upon work supported by the National Science Foundation (#1344257), the John Templeton Foundation (#48503), and the Michigan Institute for Data Science. Any opinions, findings, and conclusions or recommendations expressed in this mate-

rial are those of the author and do not necessarily reflect the views of the National Science Foundation, the John Templeton Foundation, or the Michigan Institute for Data Science.

We would like to thank Samuel Gosling for helping with the dataset collection, Shibamouli Lahiri for providing the code to calculate readability features, and Steven R. Wilson for providing the code to implement the Mairesse et al. paper that we use for prediction comparison.

References

[1] Meeker M. Internet trends 2014–Code conference. 2014;Retrieved May 28, 2014.

[2] Wendlandt L, Mihalcea R, Boyd R, Pennebaker J. Multimodal Analysis and Prediction of Latent User Dimensions. In: Proceedings of the 9th International Conference on Social Informatics (SocInfo 2017). Oxford, UK; 2017. p. 323–340.

[3] Boyd RL. Psychological text analysis in the digital humanities. In: Data analytics in digital humanities. Springer; 2017. p. 161–189.

[4] Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; 2015. p. 1–10.

[5] Conover M, Gonçalves B, Ratkiewicz J, Flammini A, Menczer F. Predicting the political alignment of Twitter users. In: Proceedings of 3rd IEEE Conference on Social Computing (SocialCom); 2011. p. 192–199.

[6] Cohen R, Ruths D. Classifying political orientation on Twitter: It's not easy! In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM 2013); 2013. p. 91–99.

[7] van der Goot R, Ljubešić N, Matroos I, Nissim M, Plank B. Bleaching Text: Abstract Features for Cross-lingual Gender Prediction. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics; 2018. p. 383–389.

[8] Ciccone G, Sultan A, Laporte L, Egyed-Zsigmond E, Alhamzeh A, Granitzer M. Stacked Gender Prediction from Tweet Texts and Images Notebook for PAN at CLEF 2018. In:

CLEF 2018 - Conference and Labs of the Evaluation; 2018. p. 11p. Available from: https://hal.archives-ouvertes.fr/hal-02013987.

[9] Mukherjee A, Liu B. Improving gender classification of blog authors. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing; 2010. p. 207–217.

[10] Rao D, Yarowsky D, Shreevats A, Gupta M. Classifying latent user attributes in Twitter. In: Proceedings of the 2nd International Workshop on Search and Mining User-generated Contents; 2010. p. 37–44.

[11] Burger JD, Henderson J, Kim G, Zarrella G. Discriminating gender on Twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing; 2011. p. 1301–1309.

[12] Van Durme B. Streaming analysis of discourse participants. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning; 2012. p. 48–58.

[13] Volkova S, Yarowsky D. Improving gender prediction of social media users via weighted annotator rationales. In: NeurIPS Workshop on Personalization; 2014. .

[14] Volkova S, Bachrach Y, Armstrong M, Sharma V. Inferring Latent User Properties from Texts Published in Social Media. In: AAAI Conference on Artificial Intelligence; 2015. p. 4296–4297.

[15] Pennacchiotti M, Popescu AM. Democrats, Republicans and Starbucks afficinados: User classification in Twitter. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2011. p. 430–438.

[16] Eisenstein J, Smith NA, Xing EP. Discovering sociolinguistic associations with structured sparsity. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1; 2011. p. 1365–1374.

[17] Rao D, Paul M, Fink C, Yarowsky D, Oates T, Coppersmith G. Hierarchical Bayesian models for latent attribute detection in social media. In: International AAAI Conference on Weblogs and Social Media; 2011. p. 598–601.

[18] Li Y, Yang L, Xu B, Wang J, Lin H. Improving User Attribute Classification with Text and Social Network Attention. Cognitive Computation. 2019 Aug;11(4):459–468. Available from: https://doi.org/10.1007/s12559-019-9624-y.

[19] Favaretto RM, Knob P, Musse SR, Vilanova F, Costa ÂB. Detecting personality and emotion traits in crowds from video sequences. Machine Vision and Applications. 2019 Jul;30(5):999–1012. Available from: https://doi.org/10.1007/s00138-018-0979-y.

[20] Al-Ghadir AI, Azmi AM. A Study of Arabic Social Media Users—Posting Behavior and Author's Gender Prediction. Cognitive Computation. 2019 Feb;11(1):71–86. Available from: https://doi.org/10.1007/s12559-018-9592-7.

[21] Favaretto RM, Knob P, Musse SR, Vilanova F, Costa ÂB. Detecting personality and emotion traits in crowds from video sequences. Machine Vision and Applications. 2019;30(5):999–1012.

[22] An G, Levitan SI, Hirschberg J, Levitan R. Deep Personality Recognition for Deception Detection. In: Interspeech; 2018. p. 421–425.

[23] Moreno DRJ, Gomez JC, Almanza-Ojeda DL, Ibarra-Manzano MA. Prediction of Personality Traits in Twitter Users with Latent Features. In: 2019 International Conference on Electronics, Communications and Computers; 2019. p. 176–181.

[24] Bose R, Dey RK, Roy S, Sarddar D. Analyzing Political Sentiment Using Twitter Data. In: Information and Communication Technology for Intelligent Systems. Springer Singapore; 2019. p. 427–436.

[25] Volkova S, Durme BV. Online Bayesian Models for Personal Analytics in Social Media. In: AAAI Conference on Artificial Intelligence; 2015. p. 2325–2331.

[26] Seabrook EM, Kern ML, Fulcher BD, Rickard NS. Predicting Depression From Language-Based Emotion Dynamics: Longitudinal Analysis of Facebook and Twitter Status Updates. J Med Internet Res. 2018 May;20(5):e168. Available from: http://www.jmir.org/2018/5/e168/.

[27] Riordan B, Wade H, Upal A. Detecting sociostructural beliefs about group status differences in online discussions. In: Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media; 2014. p. 1–6.

[28] Gottipati S, Qiu M, Yang L, Zhu F, Jiang J. An Integrated Model for User Attribute Discovery: A Case Study on Political Affiliation Identification. In: Tseng V, Ho T, Zhou ZH, Chen AP, Kao HY, editors. Advances in Knowledge Discovery and Data Mining. vol. 8443 of Lecture Notes in Computer Science. Springer International Publishing; 2014. p. 434–446.

[29] Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, et al. Personality, gender, and age in the language of social media: The open vocabulary approach. PLOS ONE. 2013 Sept;8(9):1–16.

[30] Chang J, Rosenn I, Backstrom L, Marlow C. ePluribus: Ethnicity on social networks. In: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media; 2010. p. 18–25.

[31] Mohammady E, Culotta A. Using county demographics to infer attributes of Twitter users. In: Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media; 2014. p. 7–16.

[32] Yang SH, Long B, Smola A, Sadagopan N, Zheng Z, Zha H. Like like alike: Joint friendship and interest propagation in social networks. In: Proceedings of the 20th International Conference on World Wide Web. WWW '11; 2011. p. 537–546.

[33] Gong NZ, Talwalkar A, Mackey LW, Huang L, Shin ECR, Stefanov E, et al. Predicting links and inferring attributes using a social-attribute network (SAN). In: The 6th SNA-KDD Workshop; 2012. .

[34] Filippova K. User demographics and language in an implicit social network. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL); 2012. p. 1478–1488.

[35] Nguyen D, Gravel R, Trieschnigg D, Meder T. "How old do you think I am?" A study of language and age in Twitter. In: Proceedings of the AAAI Conference on Weblogs and Social Media (ICWSM); 2013. p. 439–448.

[36] Bergsma S, Post M, Yarowsky D. Stylometric Analysis of Scientific Articles. In: Proceedings of the North American Association of Computational Linguistics. Montreal, CA; 2012. p. 327–337.

[37] Bergsma S, Dredze M, Durme BV, Wilson T, Yarowsky D. Broadly improving user classification via communication-based name and location clustering on Twitter. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2013. p. 1010–1019.

[38] Eisenstein J, O'Connor B, Smith NA, Xing EP. A Latent Variable Model for Geographic Lexical Variation. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. EMNLP '10; 2010. p. 1277–1287.

[39] Kelly EL, Conley JJ. Personality and compatibility: A prospective analysis of marital stability and marital satisfaction. Journal of Personality and Social Psychology. 1987;52(1):27.

[40] Roberts B, Kuncel N, Shiner R, Caspi A, Goldberg L. The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. Perspectives on Psychological Science. 2007;4(2):313–345.

[41] Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, Stillwell DJ, et al. Automatic Personality Assessment Through Social Media Language. Journal of Personality and Social Psychology. 2014;.

[42] Pennebaker JW, King LA. Linguistic styles: Language use as an individual difference. Journal of Personality and Social Psychology. 1999;77(6):1296.

[43] Mairesse F, Walker MA, Mehl MR, Moore RK. Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of Artificial Intelligence Research. 2007;30:457–500.

[44] Whitty MT, Doodson J, Creese S, Hodges D. A picture tells a thousand words: What Facebook and Twitter images convey about our personality. Personality and Individual Differences. 2018;133:109–114.

[45] Lay A, Ferwerda B. Predicting Users' Personality Based on Their 'Liked' Images on Instagram. In: The 23rd International on Intelligent User Interfaces, March 7-11, 2018; 2018. .

[46] Newman ML, Groom CJ, Handelman LD, Pennebaker JW. Gender differences in language use: An analysis of 14,000 text samples. Discourse Processes. 2008;45(3):211–236.

[47] You Q, Bhatia S, Sun T, Luo J. The eyes of the beholder: Gender prediction using images posted in online social networks. In: 2014 IEEE International Conference on Data Mining Workshop. IEEE; 2014. p. 1026–1030.

[48] Zhang D, Islam MM, Lu G. A review on automatic image annotation techniques. Pattern Recognition. 2012;45(1):346–362.

[49] Hossain M, Sohel F, Shiratuddin MF, Laga H. A comprehensive survey of deep learning for image captioning. ACM Computing Surveys (CSUR). 2019;51(6):118.

[50] Mithun NC, Panda R, Papalexakis EE, Roy-Chowdhury AK. Webly Supervised Joint Embedding for Cross-Modal Image-Text Retrieval. In: Proceedings of the 26th ACM International Conference on Multimedia. MM '18. New York, NY, USA: ACM; 2018. p. 1856–1864. Available from: http://doi.acm.org/10.1145/3240508.3240712.

[51] Johnson J, Karpathy A, Fei-Fei L. Densecap: Fully convolutional localization networks for dense captioning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 4565–4574.

[52] McCrae RR, John OP. An introduction to the five-factor model and its applications. Journal of Personality. 1992;60(2):175–215.

[53] John OP, Srivastava S. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of Personality: Theory and Research. 1999;2(1999):102–138.

[54] Yoder PJ, Blackford JU, Waller NG, Kim G. Enhancing power while controlling family-wise error: An illustration of the issues using electrocortical studies. Journal of Clinical and Experimental Neuropsychology. 2004;26(3):320–331.

[55] Redi M, Quercia D, Graham L, Gosling S. Like Partying? Your Face Says It All. Predicting the Ambiance of Places with Profile Pictures. In: Ninth International AAAI Conference on Web and Social Media; 2015. .

[56] Khouw N. The meaning of color for gender. Colors Matters–Research. 2002;.

[57] Van De Weijer J, Schmid C, Verbeek J, Larlus D. Learning color names for real-world applications. IEEE Transactions on Image Processing. 2009;18(7):1512–1523.

[58] Valdez P, Mehrabian A. Effects of color on emotions. Journal of Experimental Psychology: General. 1994;123(4):394.

[59] Machajdik J, Hanbury A. Affective image classification using features inspired by psychology and art theory. In: Proceedings of the 18th ACM International Conference on Multimedia. ACM; 2010. p. 83–92.

[60] Lovato P, Bicego M, Segalin C, Perina A, Sebe N, Cristani M. Faved! Biometrics: Tell me which image you like and I'll tell you who you are. IEEE Transactions on Information Forensics and Security. 2014;9(3):364–374.

[61] Gosling SD, Ko SJ, Mannarelli T, Morris ME. A room with a cue: Personality judgments based on offices and bedrooms. Journal of Personality and Social Psychology. 2002;82(3):379.

[62] Zhou B, Lapedriza A, Xiao J, Torralba A, Oliva A. Learning deep features for scene recognition using places database. In: Advances in Neural Information Processing Systems; 2014. p. 487–495.

[63] Mathias M, Benenson R, Pedersoli M, Van Gool L. Face detection without bells and whistles. In: European Conference on Computer Vision. Springer; 2014. p. 720–735.

[64] Gosling SD, Craik KH, Martin NR, Pryor MR. Material attributes of personal living spaces. Home Cultures. 2005;2(1):51–87.

[65] Fellbaum C. WordNet. Wiley Online Library; 1998.

[66] Ciaramita M, Johnson M. Supersense tagging of unknown nouns in WordNet. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics; 2003. p. 168–175.

[67] Bentivogli L, Forner P, Magnini B, Pianta E. Revising the WordNet domains hierarchy: Semantics, coverage and balancing. In: Proceedings of the Workshop on Multilingual Linguistic Ressources. Association for Computational Linguistics; 2004. p. 101–108.

[68] Finkel JR, Grenager T, Manning C. Incorporating non-local information into information extraction systems by Gibbs sampling. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics; 2005. p. 363–370.

[69] Li JJ, Nenkova A. Fast and Accurate Prediction of Sentence Specificity. In: AAAI; 2015. p. 2281–2287.

[70] Coltheart M. The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology. 1981;33(4):497–505.

[71] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems; 2013. p. 3111–3119.

[72] Oberlander J, Nowson S. Whose thumb is it anyway?: Classifying author personality from weblog text. In: COLING/ACL; 2006. p. 627–634.