How Personality Affects our Likes: Towards a Better Understanding of Actionable Images

Francesco Gelli
National University of Singapore
francesco.gelli@u.nus.edu

Xiangnan He*
National University of Singapore
xiangnanhe@gmail.com

Tao Chen
Johns Hopkins University
tchen77@jhu.edu

Tat-Seng Chua
National University of Singapore
chuats@comp.nus.edu.sg

ABSTRACT

Messages like "If You Drink Don't Drive", "Each water drop count" or "Smoking causes cancer" are often paired with visual content in order to persuade an audience to perform specific actions, such as clicking a link, retweeting a post or purchasing a product. Despite its usefulness, the current way of discovering actionable images is entirely manual and typically requires marketing experts to filter over thousands of candidate images. To help understand the audience, marketers and social scientists have been investigating for years the role of personality in personalized services by leveraging AI technologies and social network data. In this work, we analyze how personality affects user actions on images in a social network website, and which visual stimuli contained in image content influence actions from users with certain Big Five traits. In order to achieve this goal, we ground this research on psychological studies which investigate the interplay between personality and emotions. Given a public Twitter dataset containing 1.6 million user-image timeline retweet actions, we carried out two extensive statistical analysis, which show significant correlation between personality traits and affective visual concepts in image content. We then proposed a novel model that combines user personality traits and image visual concepts for the task of predicting user actions in advance. This work is the first attempt to integrate personality traits and multimedia features, and moves an important step towards building personalized systems for automatically discovering actionable multimedia content.

CCS CONCEPTS

• Information systems → Personalization; **• Human-centered computing → Collaborative and social computing**; **Social media**; **• Applied computing →** *Psychology*;

KEYWORDS

Big Five Personality; Personalized Services; Visual Sentiment

*Xiangnan He is the corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
MM'17, , October 23–27, 2017, Mountain View, CA, USA.
© 2017 Association for Computing Machinery.
ACM ISBN ISBN 978-1-4503-4906-2/17/10…$15.00
https://doi.org/10.1145/3123266.3127909

1 INTRODUCTION

Actionable content is often designed to persuade an audience to perform specific actions, such as clicking on a news article, retweeting a post or purchasing a product. Creation and discovery of personalized actionable multimedia content are performed regularly in applications such as recommender systems, advertising and education. Currently they are mostly done manually by experts who are trained to produce the most persuasive content for a target audience. For example, content may be discovered to influence people towards committing direct actions such as purchasing a product, voting for a specific candidate, or achieving indirect persuasive goals, such as educating citizens having a more considerate behaviour in public transportation. Visual content has been shown to have much more persuasive potential than text only [38, 41]. This explains why advertisements often contain rich and elaborate visual information. "*An image is worth a thousand words*" or "*Images Can't Lie*" are examples of slogans which are commonly used for educating content producers to use images for their actionable content [16].

Despite the advances in personalized services, we are still far from building a framework for automatically discovering actionable images, since the dynamics of visual persuasion are highly complex and the understanding of what are the visual stimuli that influence people's actions is limited. One of the factors that have been known for years to be important for designing effective personalized services is personality: experts would manually choose different content for an open and extrovert audience than for an introvert and neurotic one. In this paper we advance the hypothesis that the same principle applies to visual content as well, as shown in the example images in Figure 1. Nonetheless, personality has often been overlooked in personalized services, probably due to the difficulty of assessing it for a large number of users. Recently [28], psychologists found that the Big Five traits (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism) can be easily predicted from the actions performed by users in social network sites, making the problem of personality assessment far more accessible than asking subjects to fill long questionnaires.

While other works predicted personality using images [19, 35, 36], to the best of our knowledge the opposite problem of understanding how personality traits affect people's actions on images

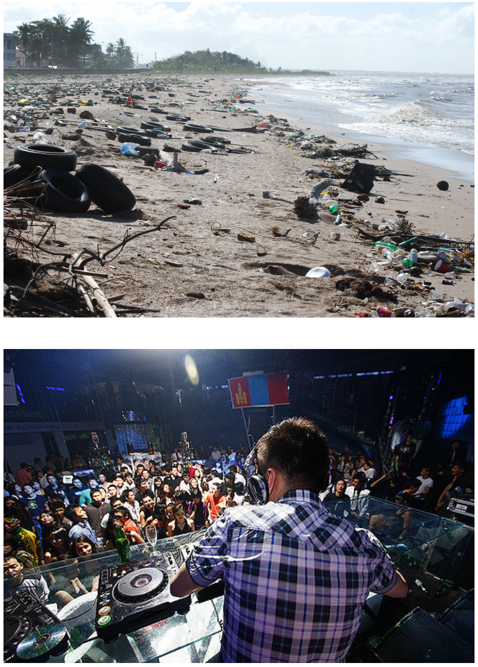

Figure 1: People with different personalities are likely to perform actions on different images. For example, conscientious people, who are usually thoughtful and have a deep sense of duty, will be probably influenced by the first picture about pollution of the environment. At the same time extroverts will mostly act on the picture below, showing a party, differently from neurotic subjects.

has not been investigated yet. In order to achieve this goal it is important to understand which visual stimuli in the image content correlate with personality traits. Psychologists and behavioral scientists studied the interplay between personality and different factors such as perception of emotions. The findings are that people with different personalities will perceive emotions differently: for example neurotic people are more prone to experience negative emotions in response to undesired events or circumstances. Recent advances in multimedia and computer vision provided the research community with effective tools for sentiment detection from images: rich visual sentiment ontologies have been developed containing thousands of affective concepts which are linked to basic emotions such as anger and serenity. The interplay of user personality and affective concepts in image content will thus be a good starting point to understand the effect of visual content on users' actions.

The huge amount of data from social network sites offers a valuable source for the study of many personalized services, since millions of users constantly interact with user generated items performing actions such as like, rate, comment and share. For this work, we constructed a public dataset containing millions of retweet actions that Twitter users performed on images inside the social network.

This work aims to answer three research questions. First, how does personality play a role in explaining users' actions on images from social network sites? Second, can we identify some of the visual stimuli that influence users with specific traits? Third, can personality help to predict users' actions in advance? To answer the first two questions, we conducted statistical studies and found significant correlation between personality Big Five traits and some affective concepts in images they retweeted. The top correlated affective concepts, either positively or negatively, can be regarded as visual stimuli that will encourage users with a specific personality fingerprint to perform actions on the images. The last question involved the design of an action prediction framework that leverages Big Five traits and affective concepts. The action prediction task involves predicting whether a user will act on a specific item in the future. Even though many effective methods have been developed to model user-item interactions in static settings of collaborative filtering [23], this problem is partially very challenging in social network sites due to the high data sparsity and dynamic items. Inspired by the state-of-the-art recommendation model Factorization Machines [22, 33], we developed a novel method named Content-Aware Factorization Machines (CAFM) that is able to model the sparse interactions between users and items while taking into account dense side information such as personality traits and image distribution. CAFM can be used to model the pairwise interaction between each user, item, personality and concept in a latent space. Experimental results demonstrate that our model is able to outperform the state-of-the-art systems designed for the task of image tweet recommendation, indicating the importance of incorporating personality in designing personalized services and discovering actionable images for a target audience.

The rest of paper is organized as follows: Section 2 gives an overview of the related work related to personality, emotions and personalized services. Secondly, section 3 presents the dataset. Sections 4 and 5 describe two statistical studies that respectively analyze personality traits separately and within an unified framework. In Section 6 we investigate the role of the Big Five traits in relation to sentiment intensity and polarity. Finally, in Section 7 CAFM is proposed and performances are evaluated against a set of baselines, followed by the conclusions and discussions of future work.

2 RELATED WORK

2.1 Personality Traits, Social Networks and Emotions

The problem of understanding the personality of individuals has been an active research topic in psychology and consumer behaviour for decades. The widely used models to represent personality include MBTI [31] and Big Five [18]. MBTI comprehends sixteen types of personality that individuals may be assigned to, while Big Five uses a combination of five orthogonal traits of: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. Briefly, *Openness* is a measure of a person's imagination and creativity. High scores in this trait often correspond to a personality that is more open to adventure, challenges and new things. The trait *Conscientiousness* reflects thoughtfulness, self-discipline and strife for achievement. High scores suggest a preference for planning than for improvisation. *Extraversion* is related to the engagement with the external world. Extrovert people are more outgoing and talkative. People who rate high on *Agreeableness* are cooperative, altruistic and emphatic. Selfish and manipulative subjects instead have low scores for this trait. Finally, *Neuroticism* is a tendency to experience negative emotions and feelings. Neurotic people tend to be moody and are stressed easily.

Since personality traits are assessed using questionnaires containing a variable size of items, several methods have been proposed to predict personality from other sources, such as natural language text. One of the most popular tools for assessing psychological features from text is LIWC: Linguistic Inquiry and Word Count [39]. The tool counts the occurrences of words that fall in one or more categories corresponding to personality and psychological states.

More recently, Kosinski et al. designed a simple logistic regression model for assessing the Big Five traits from Facebook likes [28]. The work was later extended by Quercia et al. who proposed a detector for personality traits from textual tweets [32], confirming the strong link between personality and activity in social networks. Nonetheless, there is no study that link personality to visual content.

Given the higher accessibility, user personality is becoming more commonly used in personalized services to tackle the problems of cold start and data sparsity. Elahi et al [11] adopted Big Five personality traits for recommending new places of interests using active learning. Fernandez-Tobias et al. used personality features to effectively recommend movies, music and books to new users [14]. Personality was also shown to improve personalized services in terms of recommendation diversity [5, 15].

However, none of these works considers the relations between personality and content of images in social network. [19, 35, 36] adopted image features to predict user personality from likes or timeline pictures, while [2, 12] suggest to employ multiple sources. However, the opposite task of inferring actions from personality remains unaddressed.

Several related works analyzed instead the interplay between personality and other entities that can be detected in images. A first group of works advanced the hypothesis that personality traits affect how people perceive emotions [34, 40]. They observed, for example, that extrovert subjects were more inclined to positive emotions, while neuroticism is instead correlated with negative sentiments. Diener et al. [9] found that personality is one of the factors that explain a significant amount of the variability of people's emotional and cognitive evaluations of their lives. More recently Komulainen et al. [27] utilized linear models to analyze the effect of the personality traits on daily emotional processes and found correlation between specific traits and emotional processes, like reaction to stressors or daily incidents. For example, neurotic people tend to have higher negative emotion processes in response to incidences as compared to conscientious subjects. Finally Coleman and Wu [8] included personality as one of the factors that trigger affective evaluation by voters in political elections. These works inspired us in making use of personality and affective concepts from image content to bridge the gap between images and users who acted on them in a social network.

2.2 Visual Emotion

The detection of emotional content of images is an important topic in multimedia. A large number of works has tried to classify images into one categorical emotion state such as *anger* or *amusement*. Jia et al. [24] used a factor graph semi-supervised method to infer affects from Flickr images . Principles-of-art features like symmetry and harmony were introduced by Zhao et al [42]. Borth et al [1] designed a Visual Ontology consisting of a set of more than 3,000 Adjective Noun Pairs (ANP), where each one is a concept which is strongly linked to an emotional state. ANPs are mined from the Web using keywords corresponding to the basic emotions of The Plutchik wheel of sentiments and are assigned a score in the $[-2, 2]$ range, according to the sentiment polarity. Examples or such affective concepts are *AngryFace* and *CalmSea*. ANPs have also been used to understand the link between visual sentiment and popularity of images on Flickr [17]. Jou et al. [25] extended the ontology to a larger number of concepts covering different languages and released a set of detectors based on convolutional neural networks. Given the rich ontology of emotional concepts, we plan to use the English version of this set of detectors for representing the potential emotions which are aroused by images. Each concept detected in images will be a potential emotional stimulus that may influence the actions from users with certain personality.

2.3 Personalized Services

The investigation into factors that partially explain which items will trigger people's actions has attracted vast interest. However, only a small number of these works have focused on social network items or visual content. In this section we will review separately these two groups of works.

An effective strategy to predict which users in a social network will act on an item is to adopt diffusion models, which study the initial diffusion effect of the item in the network from the time it was initially posted [7, 26]. These works use local structure of the social network and temporal features, but often ignore the content itself because of its inferior predictive power and difficulty to generalize [37]. Even though the diffusion models could achieve high performances, they cannot be applied outside social networks, since they rely heavily on social features. We are instead interested in results which can be generalized outside the context of social network structures.

Other works extend the Collaborative Filtering (CF) techniques for predicting user actions [4, 13]. Since CF approaches are known to suffer from item cold dart problem [6], these works often look into the content of the items, such as textual words or hashtags. A limitation is that only simple text features are used as content, and none of these work has looked into multimedia content.

Partially inspired from recent advances in computer vision and deep learning technologies [23], a few hybrid architectures have been proposed to recommend images. Chen et al [3] introduced a generic architecture, which learns rich item latent vectors from implicit user-item feedback and other existing information of the items, for example visual knowledge. Lei et al. [29] proposed a similar framework for pairwise personalized image ranking. The model comprehends two convolutional neural networks for capturing visual information and a fully connected network for discriminating

Table 1: Dataset Statistics

	Users	Retweets	Tweets	Ratings
Training	862	141,706	1,171,426	1,302,507
Testing		7,218	68,006	68,101

which of the two images is more likely to trigger an action from a specific user. Other works developed similar methods for the more specific case of fashion domain recommendations [20, 21].

Only a very limited number of attention has been directed to recommending images in social network scenarios, which exhibits high data sparsity and cold start problem. Liu et al. [30] introduced a deep learning model where image embeddings are learned by fusing semantic aspect and the intention of users who interacted in the past. Finally, Chen et al. [6] proposed a context-aware recommendations framework for image tweets, where context features are used, extracted both from the text and image. We consider these as state-of-the-art features for personalized image recommendations in social networks. However, the visual content is only used to the extent of extracting the image text with OCR. In our work we want to enrich the action prediction model with richer image content information and personality, in particular by leveraging the interplay between these two measures.

3 DATASET

To answer the research questions we constructed an extensive dataset with considerable number of user actions on social network images. Since our statistical studies and action prediction model are both performed on the same dataset, we describe its details and construction here[1]. In particular, we extended a publicly released dataset by Chen et al. [6] with personality and concept information. The dataset contains more than one million interactions between users and images on Twitter. The statistics of the dataset can be found in Table 1.[2] Note that the dataset contains both positive and negative actions, where the former are retweet interactions and the latter are opportunely sampled by [6]. For assessing the personality for the users in the dataset, reaching such a large number of users for a personality questionnaire is infeasible. Secondly, questionnaires are also often subject to flaws. For example, results may depend on the circumstances or mood of the subjects at the questionnaire time. Based on the well accepted theory that personality is encoded in human language, we adopted a more scalable solutions for overcoming the two problems above. We first collected a separate set of textual tweets[3] and then employed the Apply Magic Sauce API[4] [32] to extract the personality Big Five traits. According to the guidelines of the API, users with less than 200 words of Tweets were discarded, resulting in a set of 862 users. We hence represent personality of user u with a vector of five distinct values in the range of $[0, 1]$, marked as $\mathbf{p}_u$. Even if such personality prediction models have lower accuracies, we advance the hypothesis they still can help for the task of discovery of actionable media

[1]Dataset and source code are available at https://github.com/GelliFrancesco/CAFM

[2]Statistics about the dataset may not match with the original paper: since only the tweet ids were released, images have been crawled again. Image tweets which have been removed are hence missing in our version.

[3]Tweets originally written by the users were retained, retweets were discarded.

[4]Accessible on https://applymagicsauce.com/

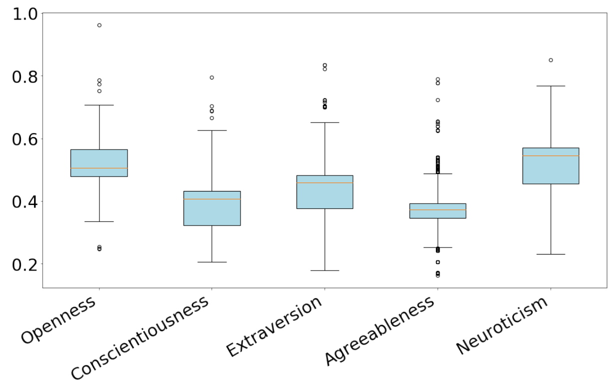

Figure 2: Box plot for personality traits. Each column represents the distribution for a single trait in the dataset, where each box contains half of the data population. Yellow lines indicate the median and outliers are marked with circles.

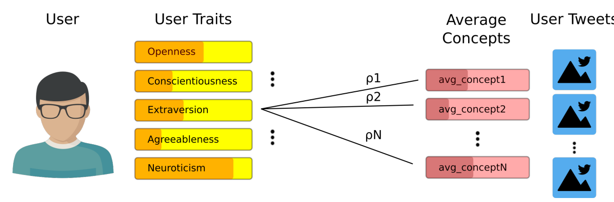

Figure 3: Single Trait Analysis: correlation is computed for each personality trait each concept average score. Average is computed among all the images retweeted by the user.

items. Figure 2 shows the distribution of the personality traits for the dataset. The figure indicates that most users have openness and neuroticism scores of higher than 0.5, while high values in conscientiousness and agreeableness are more rarely found.

We used the pretrained convolutional neural network by [25] to extract image affective features. Given a new image, this architecture is trained to compute a distribution of 4,342 affective concepts such as *SweetKiss* or *SadFace*. We denote such distribution for an image i as $\mathbf{q}_i$.

We user n and m to denote the number of users and tweets respectively. In the rest of the paper we represent the dataset with two matrices $\mathbf{P} \in \mathbb{R}^{n \times 5}$ and $\mathbf{Q} \in \mathbb{R}^{m \times 4342}$, where a row denotes respectively a user's personality vector $\mathbf{p}_u$ or an item's concept distribution $\mathbf{q}_i$. Since a user has retweeted multiple images, we define the average concept matrix as $\overline{\mathbf{Q}} \in \mathbb{R}^{n \times 4342}$, where the u-th row is computed as $\frac{1}{|Tw(u)|} \sum_{i \in Tw(u)} \mathbf{q}_i$ and $Tw(u)$ represents the set of tweets performed by user u.

4 SINGLE TRAIT STUDY

This first study analyzes the actions by considering one specific Big Five trait at a time. Note that the Big Five traits are often studied independently in psychology for the ease of interpreting the results. The goal of this study is to first show that exists significant correlation between each personality trait and some of the visual

Table 2: Most correlated image entities for each personality trait. The correlation score $\rho(t, c)$ indicates how strong the correlation is, while the sign distinguishes positive from negative correlation.

Openness		Conscientiousness		Extraversion		Agreeableness		Neuroticism	
DryRiver	0.27	Environm.Issues	0.42	TightPants	0.33	UnanimousWinner	0.35	GoodMood	0.48
HiddenPlace	0.26	FreshAir	0.40	HotSummer	0.33	OutdoorExterior	0.30	SweetKiss	0.48
OldChurch	0.25	HistoricLandscape	0.39	FitWomen	0.30	FreshAir	0.30	SadGirl	0.47
AncientChurch	0.25	YoungFarmers	0.39	BeautifulWife	0.30	PerfectLight	0.30	SweetLove	0.46
OrthodoxChristianity	0.25	RoyalBank	0.39	CrazyDays	0.29	LovelyDay	0.30	PersonalSpace	0.44
BigBrothers	-0.23	FatGirl	-0.37	ScientificResearch	-0.33	NormalLife	-0.34	LargeEvent	-0.56
ActiveKids	-0.22	SadGirl	-0.36	ScientificStudy	-0.33	PregnantWomen	-0.30	CompetitiveSport	-0.48
SpecialKids	-0.20	BadReaction	-0.35	IndustrialWaste	-0.32	HotYoga	-0.29	Presid.Candidate	-0.46
FinalGame	-0.20	SickGuy	-0.35	SlowTravel	-0.31	FunnyJokes	-0.28	Presid.Campaign	-0.46
OutdoorTeam	-0.20	SickGirl	-0.34	NaturalSelection	-0.31	SimpleMan	-0.27	FairCity	-0.46

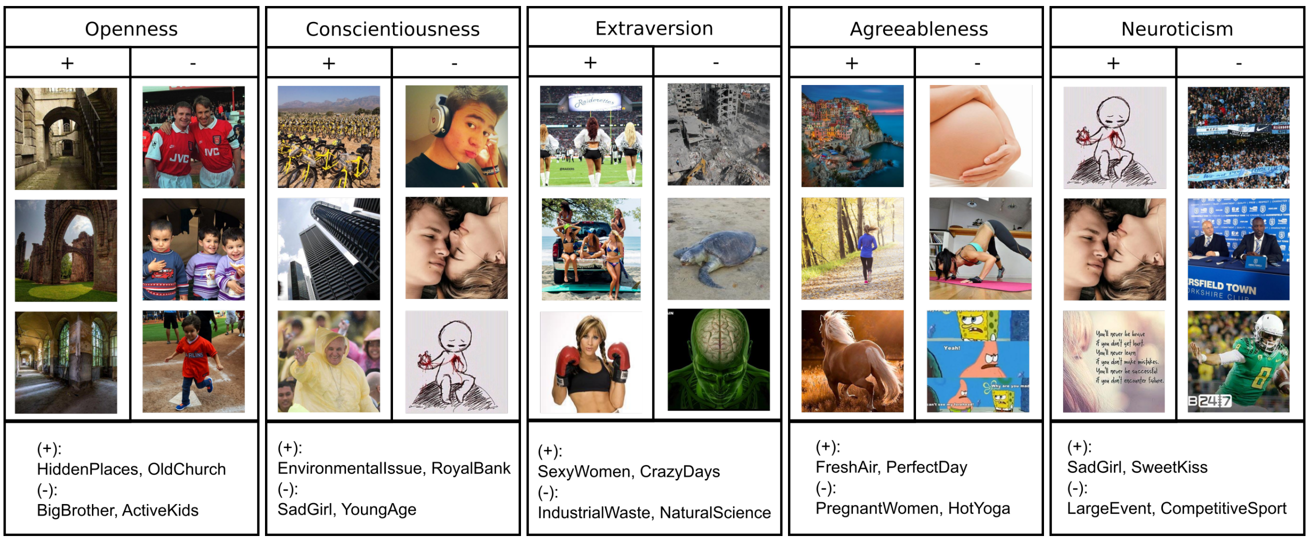

Figure 4: Visualization of Single Trait Study: for each trait six sample images are selected. The first three contain concepts which are positively correlated with the trait. Negative correlation samples are displayed in the second column. The concepts communicated by the example images are reported at the bottom.

concepts in the retweeted images, and to present a visualization of the top correlated stimuli.

For this study we retained only the retweets from the dataset, both from training and testing set, resulting in a collection of 148,924 actions from 862 users. We applied standardization of zero-mean and unit-variance to the columns of matrices $\mathbf{P}$ and $\mathbf{Q}$.

A simple overview of the single trait analysis is represented in Figure 3. For each user, all the retweeted images are collected and the concept scores are averaged, resulting in a number of observations equal to the number of users.

We use Pearson's correlation coefficient to represent how a personality trait t and an image concept c are correlated:

$$\rho(t, c) = \frac{cov(\mathbf{P}_t, \overline{\mathbf{Q}}_c)}{std(\mathbf{P}_t) std(\overline{\mathbf{Q}}_c)} \tag{1}$$

where cov is the covariance operator and std denotes the standard deviation. $\mathbf{P}_t$ and $\mathbf{Q}_c$ indicate the t-th and c-th column of $\mathbf{P}$ and $\mathbf{Q}$ respectively. A value of 1 (-1) for ρ indicates a total positive (negative) correlation, while a value of 0 indicates that the two variables have no correlation. Results were filtered with significance test, retaining only those results with a p-value lower than 0.05.

The most correlated entities for each trait are shown in Table 2. Some of the top correlated concepts are as expected, such as the case of extrovert users who appeared to perform more actions on images with concepts such as *SexyWomen*, or *CrazyDays*, while neurotic users very seldom performed actions with images involving *LargeEvent*. Notably, some of the strong correlations were unexpected, revealing new visual stimuli that may influence actions from users. For example, neurotic subjects performed many actions on images communicating sentiments such as *GoodMood*, *SweetKiss* or *SweetLove*. This effect may be due to the tendency of neurotic people to find comfort with messages that may be romantic or melancholic. Also, users with a high score on Openness were found to be influenced by images about hidden places, ancient churches and abandoned building. An explanation is that since open subjects are more creative and incline to arts, they may also be keen to exploration and mystery.

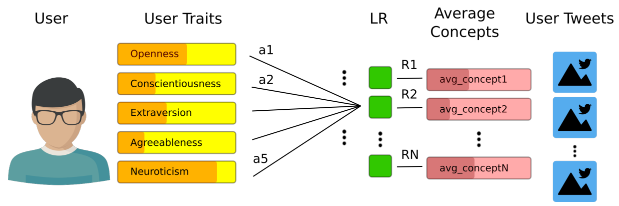

Figure 5: Multiple Correlation Analysis

Since these concepts are visual, it is important to show concrete image examples from our dataset. Some concepts and related images were selected for illustration in Figure 4. This kind of visualization helps to better understand the link between personality and visual stimuli. For example, the concepts that are positively correlated with neurotic users often correspond to cartoon-style or quote images, while this kind of stimuli often does not result in actions from conscientious users, who instead act more on the content featuring environmental issues and skyscrapers.

5 MULTIPLE CORRELATION STUDY

Studying the Big Five traits individually is a direct and straightforward analysis. However, since it considers one factor at a time, it ignores the relations between traits and may lead to misleading results. As an example, two users may have the same level of Conscientiousness, but totally different scores in Neuroticism; in this case, ignoring the latter may produce a misleading independent analysis of the former. For this reason we perform a multiple correlation study where all the traits are considered simultaneously, in order to fully understand the personality signals. Multiple correlation analysis is used to measure how well a variable can be predicted with a linear function of a set of independent variables.

In this second study we consider the Big Five traits as independent variables and image concepts as dependent variables. For each image concept we compute the coefficient of determination R^2, which indicates the proportion of the variance in the dependent variable that is explained by the independent variables. The overview of this analysis is illustrated in Figure 5. For each concept c, we learn a linear regression model from the five personality traits to predict the concept score of an user.

Let $y_{u,c}$ be the predicted value for concept c and user u and $\overline{y_c}$ be the average among all the users. The coefficient of determination is computed for each entity as follows:

$$R^2(c) = 1 - \frac{\sum_u (y_{u,c} - \overline{\mathrm{Q}}_{u,c})^2}{\sum_u (y_{u,c} - \overline{y_c})^2} \tag{2}$$

where $y_{u,c}$ is the linear regression prediction given trait-specific parameters $a_{t,c}$:

$$y_{u,c} = \sum_{t=1}^{5} a_{t,c}\mathrm{P}_{u,t} + \alpha \tag{3}$$

Since an intercept was used, the previous equation is equal to the square of the coefficient of multiple correlation $R = \sqrt{R^2}$, a commonly used metric in multiple regression analysis. Differing from previous study, this metric ranges from 0 to 1, where the higher is the score, the better is the predictability from the independent variables. In order to analyze the results, we look first into the

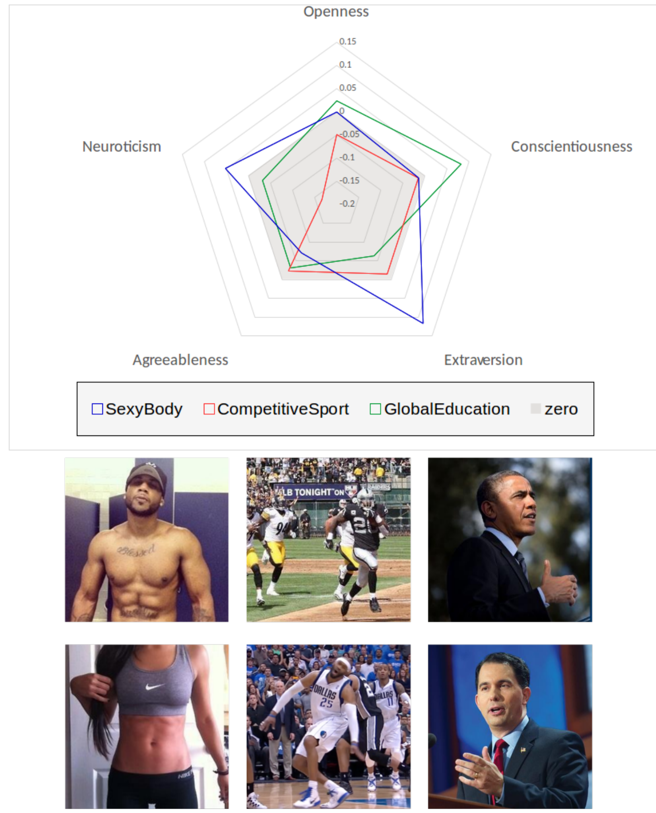

Figure 6: Three example concepts. At the top the coefficients of linear regression are shown in the radar plot.The points falling in the gray area are negative parameters, corresponding hence to traits which have a negative contribution for the concept. On the bottom three example figures from the dataset for the three concepts.

coefficient of multiple correlation R and then investigate the contribution of different traits by looking into the linear regression parameters $a_{t,c}$. The density function of R is shown in Figure 8. The distribution lies in the interval $[0.03, 0.61]$, with the mean of 0.24. Specifically, 516 out of 4,342 concepts have a multiple correlation coefficient larger than 0.4 and can be visualized in the word cloud in Figure 7. The larger a concept in the cloud, the more the personality of users will influence his or her actions on the corresponding images. We notice that many of the top concepts were also detected in the previous analysis (Table 2), such as *LargeEvent* or *EnvironmentalIssue*, but there are also new concepts such as *CleanAir*, *NaturalGas* or *GorgeousWife*.

The parameters of linear regression $a_{t,c}$ will give additional insight into which traits contribute most to the prediction. The sign discriminates between positive or negative contribution, while the absolute value indicates the strenght of the correlation In order to visualize such relations, each concept c in figure 7 is assigned different colors according to the most dominant trait for c, which is set to $\mathrm{argmax}_t |a_{t,c}|$. Different colors are then used for positive and negative sign. It is evident how neuroticism is the dominant trait for a big part of top concepts. Since in the single trait study this trait

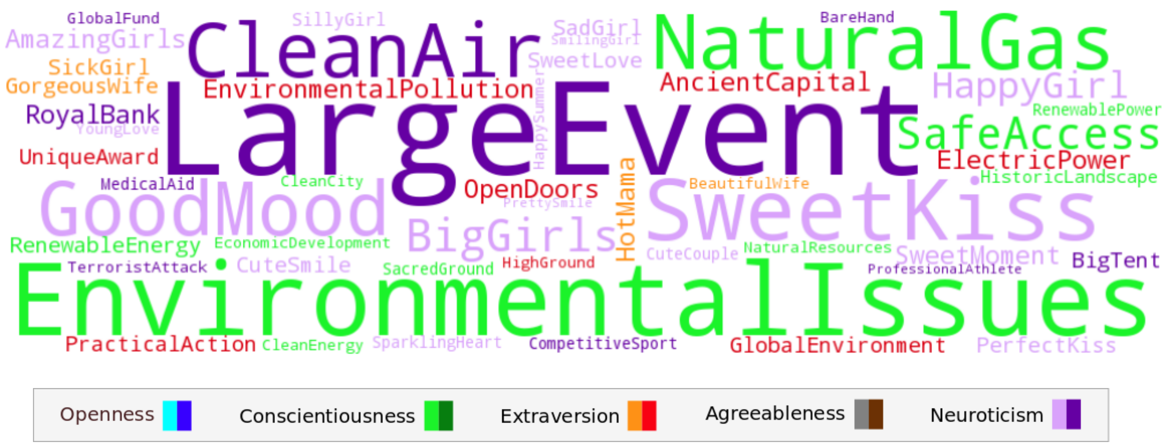

Figure 7: Word cloud representation of the multiple correlation study. Words are colored according to the dominant trait. For each trait light and dark colors indicate positive and negative correlation respectively.

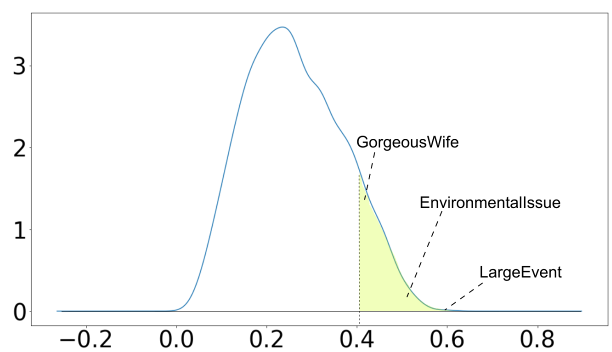

Figure 8: Density Function of the Coefficient of Multiple Correlation

has the largest absolute value correlation, we therefore conclude that neurotic users are more affected by the emotive content of images. This is not surprising since among the Big Five traits, the definition of the neuroticism trait is the most related to sentiment.

In Figure 6 the regression parameters are visualized in a radar plot for three top concepts: *SexyBody*, *CompetitiveSport* and *GlobalEducation*. This example shows how extrovert users are easily influenced by visual stimuli of partially naked bodies or parties. Among these concepts, *SexyBody* was chosen for illustration in this example. Similar results were found in [17], where 10 out of 35 of their top concepts started with the adjective *sexy*. Neurotic people also tend to act on this kind of content, but are strongly resistant to images communicating concepts related to large events and sports, such as *CompetitiveSport*. Finally, users with a high conscientiousness score are quite sensitive to the concept *GlobalEducation*, but also to other visual stimuli related to politics and environment.

6 SENTIMENT INTENSITY AND POLARITY

Users perform actions in social network websites for a variety of reasons. For example, a user may share an image because he or she strongly likes it, while another reason can be delivering some kind

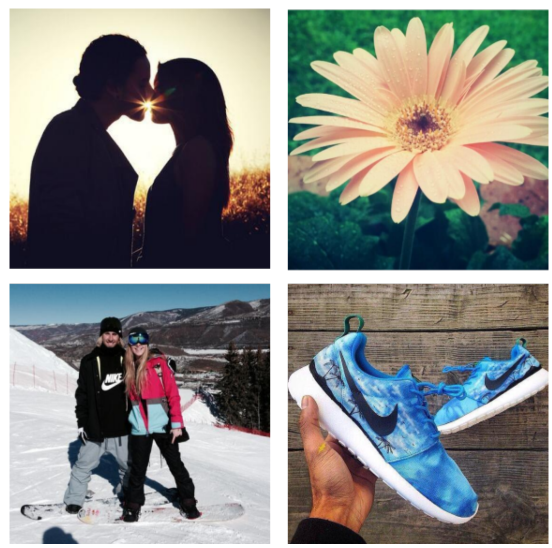

Figure 9: Emotive vs Informative Content. Users may perform actions on the two images on the top because of their emotive content, while in the two examples on the bottom the reason may be to inform the friends about a trip or the purchase of a new pair of shoes.

of information. Such user intents are often related to the nature of the content, which may be emotive or informative (Figure 9). While in our previous analyses we treated equally all the concepts in the ontology, two of the most important features of emotions are sentiment intensity and polarity. Given a content, the intensity measures how strong an emotion is, while the polarity divides the sentiment into positive, negative and neutral. We hence perform an additional investigation to understand the role of the Big Five traits in relation to sentiment intensity and polarity.

In the visual ontology we use, a score S_c is assigned to each concept c, where the sign distinguishes positive from negative emotions. As $\mathbf{q}_i$ is the concept distribution over image concepts,

Table 3: Correlation between Personality Traits and Sentiment Intensity and Polarity. Results are filtered using a significance threshold of 0.05. The more important results are marked in bold.

t	$\rho(t, \overline{\mathbf{s_{int}}})$	$\rho(t, \overline{\mathbf{s_{pol}}})$
openness	-	-0.140
conscientiousness	**-0.365**	-0.217
extraversion	0.173	**0.3014**
agreeableness	-0.139	-0.067
neuroticism	**0.513**	**0.346**

we compute sentiment intensity and polarity $\mathbf{s_{pol}}$ as follow:

$$\mathbf{s_{int}} = \sum_c Q_c |S_c| \qquad \mathbf{s_{pol}} = \sum_c Q_c S_c$$

where $\mathbf{s_{int}} \in \mathbb{R}^m$ and $\mathbf{s_{pol}} \in \mathbb{R}^m$ Analogous to previous studies, for each user we average the score over his retweets, computing $\overline{\mathbf{s_{int}}}$ and $\overline{\mathbf{s_{pol}}}$. Table 3 shows the Pearson's correlation coefficient between the two metrics and the Big Five traits. Because of the importance of negative values, data were not standardized to zero mean in this analysis. The first observation is that neurotic users are strongly affected by high intensity emotions. Secondly, it is interesting to observe that neurotisicm is also positively correlated with sentiment polarity, which reveals that positive emotions are important to stimulate actions from subjects who are commonly inclined to negative feelings. Notably, conscientiousness correlates negatively with sentiment intensity, hinting that efficient and organized people are more likely to act on neutral content, such as informative images. Finally, sentiment polarity positively correlates with extraversion, confirming that positive stimuli are effective to influence an outgoing audience.

7 ACTION PREDICTION

The significant correlation between personality and image content points towards the development of an action prediction model. The task aims to forecast in advance whether a user will perform an action on a specific image. Given a user u and an image i, the goal is to estimate the probability that an action will occur $p(act(u, i))$.

7.1 Prediction Model

This section describes the development of Content-Aware Factorization Machines (CAFM), which is a general framework for enriching sparse user-item interactions with dense content features. The motivation is that in our action prediction problem as in most hybrid recommendation systems, both sparse and dense representations are used; the former is for user-item interactions and the latter is used for content feature representation respectively. In this work we extend the popular architecture of factorization machines (FM) [22, 33] to integrate the dense features with high dimensionality, as occur frequently in multimedia due to dense image or video embeddings. CAFM is built on the flexibility of FM and can be generalized to any scenarios with both sparse and dense input. Similar to the notation of factorization machines, we represent an input data with a sparse vector $\mathbf{x_s}$, where users and items are one hot encoded. Other sparse features, such as the occurrences of words or explicit user history, can be incorporated in $\mathbf{x_s}$ without any distinction. Dense features are instead indicated as $\mathbf{x_d}$. We then represent a training instance $\mathbf{x} = (\mathbf{x_s}, \mathbf{x_d})$ with a sparse and a dense component.

For the sparse part, a traditional FM model is adopted:

$$\hat{y}_s(\mathbf{x}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} < \mathbf{v_i}, \mathbf{v_j} > \mathbf{x}_i \mathbf{x}_j + \sum_{i=1}^{N} \beta_i \mathbf{x}_i + \beta \tag{4}$$

where $\mathbf{v_i}$ and $\mathbf{v_j}$ are the latent representations for features i and j respectively. We remark that even though this formulation can also handle dense features, it will not scale with large dense dimensionality, since it considers the interaction between each pair of dense component as well.

All dense features can be incorporated in the second part of the model instead. These can be grouped into C independent features or concatenated into a single dense feature vector. The dense component has the form:

$$\hat{y}_d(\mathbf{x}) = \sum_{c1=1}^{C} \sum_{c2=c1+1}^{C} < \mathbf{W_{c1}f_{c1}} + \boldsymbol{\beta}_{c1}, \mathbf{W_{c2}f_{c2}} + \boldsymbol{\beta}_{c2} > \tag{5}$$

This component models the interactions between the C groups of dense features, which are mapped to the latent space with a linear operator. Finally, interactions between the n sparse and dense features are modeled with the following component:

$$\hat{y}_{s,d}(\mathbf{x}) = \sum_{c=1}^{C} \sum_{i=1}^{N} < \mathbf{v_i}, \mathbf{W_c f_c} + \boldsymbol{\beta}_c > \mathbf{x}_i \tag{6}$$

The final model is the summations of the three components followed by sigmoid operator:

$$\hat{y}(\mathbf{x}) = \frac{1}{1 + e^{-(\hat{y}_s(\mathbf{x}) + \hat{y}_d(\mathbf{x}) + \hat{y}_{s,d}(\mathbf{x}))}} \tag{7}$$

The described model learns the interaction among all pairs of features similar to factorization machines and also scale with high dimensionality dense input. The reason is that an embedding is not computed for each component of the dense input, but for each of the C groups. In other words, if M is the size of the dense input, the model learns $N + C$ embedding, when traditional FM would have learned $N + M$, thus saving significant amount of computation.

7.2 Results

We here present the results obtained with CAFM on the Twitter image dataset. For this task the negative samples in the Twitter dataset were also used, resulting in more than 1.3 million training and 68,000 testing samples. Among the sparse features, we use one-hot encoder for items and users and employ the state-of-the-art contextual text features for image tweet recommendation [6]. Such features are occurrences of words that may occur directly in the tweet, contained in the image or in a linked website. Two dense features groups were employed, namely the Big Five user personality traits and the distribution of 4,342 image concepts.

An illustration of how the architecture of the model is applied to this task is given in Figure 10. A latent embedding is learned for each user, item and context feature. For an input entry x, embeddings will be selected according to the sparse input. Personality and concept embeddings are instead computed with a linear layer from the dense features. Interactions between the selected embeddings will then

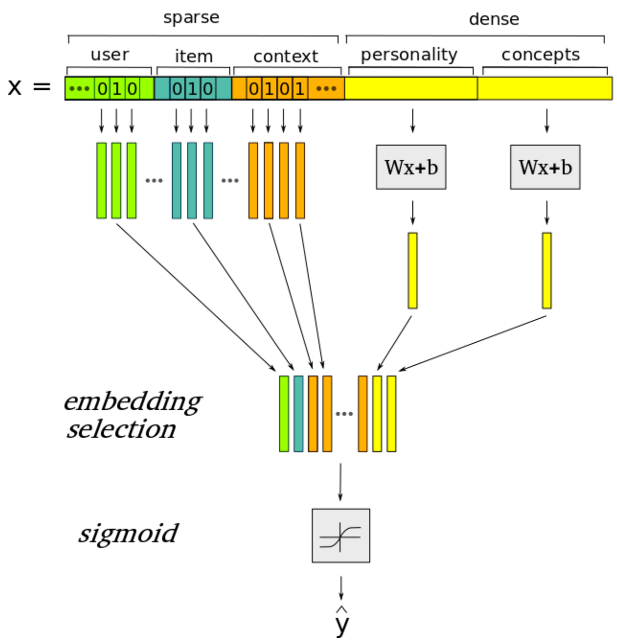

Figure 10: Model Architecture of CAFM

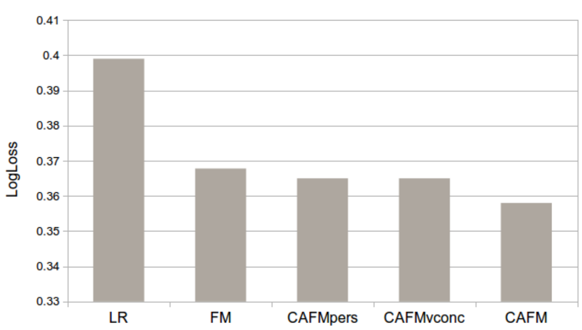

Figure 11: Performances of CAFM in terms of log loss

be computed according to equation 7 and a sigmoid operator will yield the prediction $p(act(u, i))$.

The state-of-the-art performances on this dataset were obtained by CITING [6] using context features. Since their task is a ranking problem, where the goal is to provide a personalized list of items for a user, we cannot compare directly with their method. Instead, we employ the same context features used by the authors with other methods commonly used in classification problems, that is, the combination of linear regression (LR) and factorization machines (FM). We used a similar experimental setting for CAFM and the baselines as well. Hyper-parameters were tuned separately for different models using grid search. For CAFM, we used a learning rate of 0.001 with 64 hidden dimensions, and optimized the logarithmic loss with the AdaGrad optimizer [10]. We initialized weights for feature embeddings and linear layers with a truncated Gaussian distribution. Since approximately 90% of the testing set comprehends negative samples, we sampled each training batch by selecting 25% positive instances and 75% negative ones to reduce the class imbalance.

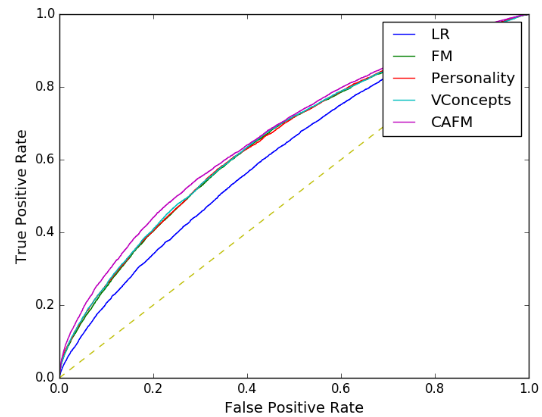

Figure 12: ROC Curve. The higher the curve, the more accurate is the classifier.

Table 4: Performance of CAFM in terms of ROC AUC

LR	FM	$CAFM_{pers}$	$CAFM_{vconc}$	$CAFM$
0.618	0.656	0.658	0.658	**0.673**

Figure 11 presents the results with respect to the logarithmic loss. Only personality traits and visual concept features are used for $CAFM_{pers}$ and $CAFM_{vconc}$ respectively, while both concepts and personality are included in $CAFM$. We first observe that $CAFM_{pers}$ achieves lower loss compared with FM, confirming that user personality assessed with online prediction models is accurate enough to improve prediction results. As our CAFM jointly models user personality traits and image concepts, it achieves the best performance as compared with the other methods and the individual features. The classification performance is shown with the ROC curve in Figure 12, which plots the true positive rate against the false positive rate. Table 4 reports the area under the ROC curve: the probability that a retweet will be given a rank higher than a negative sample is 0.673. From these results we can conclude that personality and image concepts don't help improving the performances from the FM baseline significantly when used independently; however they contribute boosting the performance when employed together.

8 CONCLUSIONS AND FUTURE WORK

This work is a first step towards automatically discovering actionable images for users according to their personality. On a dataset with hundreds of thousands of image retweets, the Big Five personality traits are found to be significantly correlated with a set of affective visual concepts, underlying how the presence or absence of such visual stimuli in images makes an item more or less actionable for individuals with specific personality fingerprints. Visualization of statistical studies showed some expected correlation that experts may already be aware of and also new insights, such as positive visual concepts such as *GoodMood* or *SweetKiss* being actionable for neurotic users. A new Content Aware Factorization Machines model produced superior performances as compared to state-of-the-art methods for the task of action prediction, building the basic block toward automatic discovery of actionable images. Among the future works, we plan to use personality information to provide personalized discovery of what images are most actionable for a specific mass educative messages such as "*start eating healthy now*" or "*stop violence against women*". This research direction will hence be of critical importance in applications where the right multimedia content may make a difference.

ACKNOWLEDGMENTS

NExT research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IRC@SG Funding Initiative.

REFERENCES

[1] Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. 2013. Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs. In *Proceedings of the 21st ACM International Conference on Multimedia*. ACM, 223–232.

[2] Kseniya Buraya, Aleksandr Farseev, Andrey Filchenkov, and Tat-Seng Chua. 2017. Towards User Personality Profiling from Multiple Social Networks. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. AAAI.

[3] Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, and Tat-Seng Chua. 2017. Attentive Collaborative Filtering: Multimedia Recommendation with Item-and Component-Level Attention. In *Proceedings of the 40th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 335–344.

[4] Kailong Chen, Tianqi Chen, Guoqing Zheng, Ou Jin, Enpeng Yao, and Yong Yu. 2012. Collaborative Personalized Tweet Recommendation. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 661–670.

[5] Li Chen, Wen Wu, and Liang He. 2016. *Personality and Recommendation Diversity*. Springer International Publishing, 201–225.

[6] Tao Chen, Xiangnan He, and Min-Yen Kan. 2016. Context-aware Image Tweet Modelling and Recommendation. In *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 1018–1027.

[7] Justin Cheng, Lada Adamic, P. Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. 2014. Can Cascades Be Predicted?. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 925–936.

[8] Renita Colemann and Denis Wu. 2015. *Image and Emotion in Voter Decisions*. Lexington Books.

[9] Ed Diener, Shigehiro Oishi, and Richard E Lucas. 2003. Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual review of psychology* 54, 1 (2003), 403–425.

[10] John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research* 12, Jul (2011), 2121–2159.

[11] Mehdi Elahi, Matthias Braunhofer, Francesco Ricci, and Marko Tkalcic. 2013. Personality-based active learning for collaborative filtering recommender systems. In *Congress of the Italian Association for Artificial Intelligence*. Springer International Publishing, 360–371.

[12] Aleksandr Farseev, Liqiang Nie, Mohammad Akbari, and Tat-Seng Chua. 2015. Harvesting multiple sources for user profile learning: a big data study. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*. ACM, 235–242.

[13] Wei Feng and Jianyong Wang. 2013. Retweet or Not?: Personalized Tweet Reranking. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*. ACM, 577–586.

[14] Ignacio Fernández-Tobías, Matthias Braunhofer, Mehdi Elahi, Francesco Ricci, and Iván Cantador. 2016. Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction* 26, 2 (2016), 221–255.

[15] Bruce Ferwerda, Mark P Graus, Andreu Vall, Marko Tkalcic, and Markus Schedl. 2016. The Influence of Users' Personality Traits on Satisfaction and Attractiveness of Diversified Recommendation Lists.. In *EMPIRE@ RecSys*. 43–47.

[16] Robert H Gass and John S Seiter. 2015. *Persuasion: Social influence and compliance gaining*. Routledge.

[17] Francesco Gelli, Tiberio Uricchio, Marco Bertini, Alberto Del Bimbo, and Shih-Fu Chang. 2015. Image Popularity Prediction in Social Media Using Sentiment and Context Features. In *Proceedings of the 23rd ACM International Conference on Multimedia*. ACM, 907–910.

[18] Lewis R Goldberg. 1990. An alternative "Description of personality": The Big-Five factor structure. *Journal of personality and social psychology* 59, 6 (1990), 1216.

[19] Sharath Chandra Guntuku, Weisi Lin, Jordan Carpenter, Wee Keong Ng, Lyle Ungar, and Daniel Preotiuc-Pietro. 2017. Studying Personality through the Content of Posted and Liked Images on Twitter, In WebSci'17 Proceedings. *Web Science*.

[20] Ruining He, Chunbin Lin, and Julian McAuley. 2016. Fashionista: A Fashion-aware Graphical System for Exploring Visually Similar Items. In *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 199–202.

[21] Ruining He and Julian McAuley. 2016. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI Press, 144–150.

[22] Xiangnan He and Tat-Seng Chua. 2017. Neural Factorization Machines for Sparse Predictive Analytics. In *Proceedings of the 40th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 355–364.

[23] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *Proceedings of the 26th international conference on World Wide Web*. http://www.comp.nus.edu.sg/˜xiangnan/papers/ncf.pdf, 173–182.

[24] Jia Jia, Sen Wu, Xiaohui Wang, Peiyun Hu, Lianhong Cai, and Jie Tang. 2012. Can We Understand Van Gogh's Mood?: Learning to Infer Affects from Images in Social Networks. In *Proceedings of the 20th ACM International Conference on Multimedia*. ACM, 857–860.

[25] Brendan Jou, Tao Chen, Nikolaos Pappas, Miriam Redi, Mercan Topkara, and Shih-Fu Chang. 2015. Visual Affect Around the World: A Large-scale Multilingual Visual Sentiment Ontology. In *Proceedings of the 23rd ACM International Conference on Multimedia*. ACM, 159–168.

[26] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 137–146.

[27] Emma Komulainen, Katarina Meskanen, Jari Lipsanen, Jari Marko Lahti, Pekka JylhÃď, Tarja Melartin, Marieke Wichers, Erkki IsometsÃď, and Jesper Ekelund. 2014. The Effect of Personality on Daily Life Emotional Processes. *PLOS ONE* 9, 10 (2014), 1–9.

[28] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110, 15 (2013), 5802–5805.

[29] Chenyi Lei, Dong Liu, Weiping Li, Zheng-Jun Zha, and Houqiang Li. 2016. Comparative Deep Learning of Hybrid Representations for Image Recommendations. *CoRR* abs/1604.01252 (2016).

[30] Shaowei Liu, Peng Cui, Wenwu Zhu, and Shiqiang Yang. 2015. Learning Socially Embedded Visual Representation from Scratch. In *Proceedings of the 23rd ACM International Conference on Multimedia*. ACM, 109–118.

[31] Peter B. Myers Myers, Isabel Briggs. 1995. *Gifts Differing: Understanding Personality Type*. CPP, Inc.

[32] Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. 2011. Our twitter profiles, our selves: Predicting personality with twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom)*. IEEE, 180–185.

[33] Steffen Rendle. 2010. Factorization machines. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 995–1000.

[34] Cheryl L Rusting. 1998. Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. *Psychological bulletin* 124, 2 (1998), 165.

[35] Jitao Sang, Huaiwen Zhang, and Changsheng Xu. 2016. Visual BFI: An Exploratory Study for Image-Based Personality Test. In *Pacific Rim Conference on Multimedia*. Springer International Publishing, 95–106.

[36] Cristina Segalin, Dong Seon Cheng, and Marco Cristani. 2017. Social profiling through image understanding: Personality inference using convolutional neural networks. *Computer Vision and Image Understanding* 156 (2017), 34 – 50.

[37] Benjamin Shulman, Amit Sharma, and Dan Cosley. 2016. Predictability of Popularity: Gaps between Prediction and Understanding. *CoRR* abs/1603.09436 (2016). http://arxiv.org/abs/1603.09436

[38] Georg Stenberg. 2006. Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology* 18, 6 (2006), 813–847.

[39] Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29, 1 (2010), 24–54.

[40] Auke Tellegen. 1985. Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. (1985).

[41] Douglas Rudy Vogel, Gary W Dickson, John A Lehman, and others. 1986. Persuasion and the role of visual presentation support: The UM/3M study. (1986).

[42] Sicheng Zhao, Yue Gao, Xiaolei Jiang, Hongxun Yao, Tat-Seng Chua, and Xiaoshuai Sun. 2014. Exploring Principles-of-Art Features For Image Emotion Recognition. In *Proceedings of the 22Nd ACM International Conference on Multimedia*. ACM, 47–56.