YOU ARE WHAT YOU TWEET...PIC! GENDER PREDICTION BASED ON SEMANTIC ANALYSIS OF SOCIAL MEDIA IMAGES

Michele Merler, Liangliang Cao, John R. Smith

IBM TJ Watson
Yorktown Heights, NY
{mimerler,liangliang.cao,jsmith}@us.ibm.com

ABSTRACT

We propose a method to extract user attributes from the pictures posted in social media feeds, specifically gender information. While traditional approaches rely on text analysis or exploit visual information only from the user profile picture or colors, we propose to look at the distribution of *semantics* in the pictures coming from the whole feed of a person to estimate gender. In order to compute such semantic distribution, we trained models from existing visual taxonomies to recognize objects, scenes and activities, and applied them to the images in each user's feed. Experiments conducted on a set of ten thousand twitter users and their collection of half a million images revealed that the gender signal can indeed be extracted from the users image feed (75.6% accuracy). Furthermore, the combination of visual cues resulted almost as strong as textual analysis in predicting gender, while providing complementary information that can be employed to further boost gender prediction accuracy to 88% when combined with textual data. As a byproduct of our investigation, we were also able to extrapolate the semantic categories of posted pictures mostly correlated to males and females.

Index Terms— Gender Prediction, Visual Analytics, Social Media, Multimodal Information Extraction

1. INTRODUCTION AND RELATED WORK

Gender prediction from social media profiles has attracted great interest in recent years. While in certain cases such information can be explicitly provided by the user, for the vast majority of cases it remains unknown. A great body of work has focused on estimating gender from textual analysis of diverse sources such as tweets [1], hashtags[1], psycho-linguistic features [2], conceptual attributes [3], topic modeling on Pinterest boards names [4], and first name analysis [5]. While textual analysis has proved quite powerful, it is not perfect [6] and suffers from the need to develop language specific models [7] for different cultures/nationalities.

Some methods tried to alleviate such shortcoming by making use of a user's network analysis, inferring gender from friends connections [8, 9] and even from the list of celebrities that one follows [10].

Fig. 1. Examples of problematic cases for gender estimation from profile pictures: (a) occlusion, (b) face not visible, (c) celebrity picture, (d) multiple people and (e) pictures not portraying people.

So far, visual information has only been marginally considered for gender prediction. Alowibdi et al.[11] analyze the colors adopted by users in their profiles. The user profile picture represents an obvious choice to extract gender information from a social media profile, and even the mere fact that a user shares a profile picture can be indicative [12]. A large portion of profiles contain a clear view of the face of a user, therefore using state of the art face gender recognition methods constitutes a powerful gender prediction cue. However, as exemplified in Figure 1 and demonstrated in Section 5, profile picture face analysis alone is not sufficient for fully reliable gender estimation. Besides the technical challenges posed by out of focus, perspective distortions and occlusions (1-a), in many cases the face of the user might not even be visible, due to specific pose choices (1-b). Furthermore, some users adopt a picture of their favorite celebrity (1-c), or pictures with two people or group photos containing individuals of different gender (1-d). Even most interestingly, users post photos which don't even contain humans as their profile(1-e). We claim that such pictures carry meaningful insights about the users' interests and attributes, which are in turn correlated to gender.

[1] `http://totems.co/blog/machine-learning-nodejs-gender-instagram`

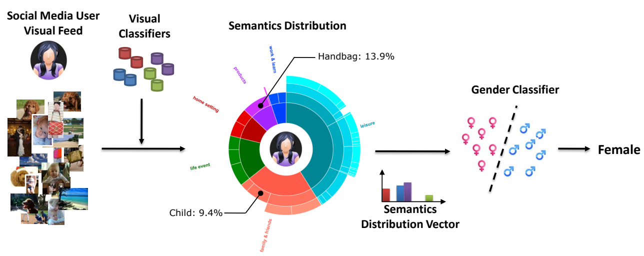


Fig. 2. Proposed gender classification pipeline.

To the best of our knowledge, Ma et al.[13] have been the only ones to use image analytics from social media streams to estimate users' gender. However, their approach is limited to analyzing only the images from the user feed, excluding non-face profile pictures. The authors employed a restricted set of classifiers which were built ad-hoc for a small dataset of a few hundred Twitter users. Furthermore, they did not explore the combination of visual and textual analysis.

Following the same logic, we apply a set of visual semantic classifiers to the entire collection of images and videos in a user's feed, and train a gender predictor on top of the aggregated semantic scores from such classifiers across each user's collection. Our approach is detailed in Figure 2: first we collect all the images from a user's social media feed, we then extract a vector containing the distribution of aggregated responses of a set of visual classifiers across all the images, and finally we learn a gender predictor on top of it.

Experiments conducted on a set of ten thousand twitter users and their collection of half a million images revealed that the gender signal can indeed be extracted from the users image feed (75.6% accuracy). Furthermore, the combination of visual cues resulted almost as strong as textual analysis in predicting gender, while providing complementary information that can be employed to further boost gender prediction accuracy to 88% when combined with textual data. As a byproduct of our investigation, we were also able to extrapolate the semantic categories of posted pictures mostly correlated to males and females.

The remainder of the paper is organized as follows: Section 2 introduces the details of our visual analytics pipeline for gender prediction, we comment on Textual Analytics and multimodal information fusion in Sections 3 and 4, respectively. We presents the experimental settings and results in Section 5, and draw conclusions in Section 6.

2. VISUAL INFORMATION EXTRACTION

In this Section we introduce our framework to estimate user gender from visual information in social media profiles, using three different sources: profile pictures, images from the feed, and profile color patterns.

2.1. Profile Picture Analysis

We adopted a two-channels analysis approach to users profile pictures: in one channel we applied a state of the art face detector and face analysis based gender estimator, while in the other we performed an analysis on top of a set of general concept classifiers, similarly to what we did for all the pictures from the user's feed.

2.1.1. Face Based Gender Predictor

In order to perform face based gender estimation, we adopted the free api from the commercial system Face++[2], which employs state of the art face detection, salient points identification, registration and attributes extraction algorithms including gender, age, facial expressions and accessories (glasses, hats, etc.). For each input image, the system returns the detected faces together with their attributes and confidence scores in a scale from 0 to 100. We refer to their work for details of the system [14]. In the dataset of 10K Twitter profiles we analyzed in the experiments, only in 54.81% of the cases the system detected a single face. Including the cases where multiple faces where detected with a majority of one gender represented, amounted to 58.71% of the users. When more than one face was detected, we predicted gender by majority voting, or by confidence score in case an equal number of male and female faces where found.

2.1.2. Profile Picture Semantics

We claim that even in cases where a user's face is not portrayed in his/her profile picture, the choice of subject for such picture is correlated with the user's gender.

We therefore employed a set of visual classifiers to recognize the content of those images and used their predictions as a feature to estimate users' gender. The choice of which categories to recognize in the pictures is not trivial. While we suspect that a set of visual classifiers specifically tailored and trained on the dataset used in this work's experiments would have provided better performance, we tried to re-use a subset of pre-existing classifiers which had been trained in the context of event detection from video collections [15]. One main motivation behind this choice was to generate a set of concepts which could be re-used for other datasets and not overly specific to the one inspected in this work.

We chose the following 25 categories: *Adult, Animal, Baby, Beach, Boy, Brand Logo, Building, CGI, Car, Cat, Child, Dog, Elderly Man, Elderly Person, Elderly Woman, Female Adult, Girl, Horse, Human Portrait View, Human, Icon, Male Adult, Motorcycle, Nature, Two People*.

In order to qualitatively evaluate our choice of visual classifiers and determine the most discriminative ones for gender, we trained two linear SVMs on top of the Semantic Model

[2]www.faceplusplus.com

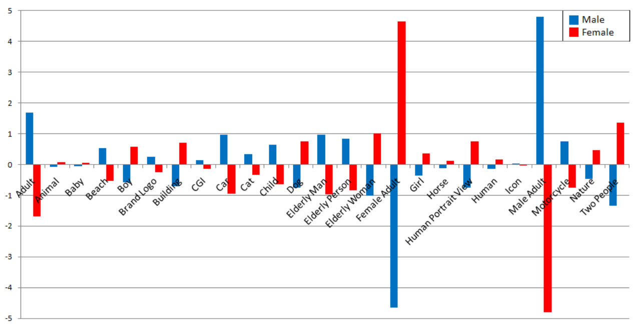


Fig. 3. Weights of the most discriminative categories correlated to profile pictures.

Vector: one using the male user profile pics as positives and the female ones as negatives, and the other inverting the roles.

In Figure 3 are reported the weights of the SMVs, male in blue and female in red. Many weights confirm obvious intuitions (for example *Male Adult* with a large positive weight for male, and a large negative weight for female). Some are more interesting, for example male users seem to be *Cat* lovers, whereas female users seem to prefer *Dog*. Male users post more vehicles (*Car* and *Motorcycle*) while female users have more profile pictures with friends (Two people) and landscapes, both rural (*Nature*) and urban (*Building*).

2.2. Feed Pictures Analysis

We apply the same type of semantic analysis described in the previous Section also to all the images in a users feed.

We employed three sets of semantic visual classifiers, and tested their performance as gender prediction signals alone or in combination. The list of semantics were chosen among the ones developed to recognize visual events from consumer videos. Details of the classifier training procedures and categories can be found in [15], while the full lists can be found online[3].

SMV 51 : a set of 51 classifiers trained as ensemble SVMs on top of standard visual descriptors, using images crawled from web search engines as training data. This initial set was chosen to be compact (for efficiency purposes) yet descriptive, trying to cover topics that people traditionally share on social media such as sports, life events, products, home related, pets, etc. For each image, we obtain a semantic model vector (SMV) of 51 concatenated prediction scores, one for each visual model.

SMV 717 : same as SMV 51, but with an extended set of 717 categories.

SMV Deep1000 : a set of 1K classifiers trained from ImageNet using a convolutional deep neural network, extracted using the Caffe package [4].

[3]http://www.cs.columbia.edu/~mmerler/gender/

[4]http://caffe.berkeleyvision.org/

While the approach is similar to the profile picture analysis, in this context we are looking at a *collection* of multiple images. Therefore the assumption is that the *distribution* of categories depicted in the images posted by a user is correlated to his/her gender. In fact, for each concept C_i, we have not one but a *set* of scores $C_i(x_j)$, with $j = 1, ..., N_k$ where N is the number of images posted by a user k.

We therefore tested different approaches to feed such distribution to the final gender classifier. As baselines we adopted standard pooling operations (max, average, average of the top quartile)

$$C_i(k) = pooling\left(C_i(x_j)\right) \tag{1}$$

We also tested a count based approach, where we counted the number of pictures in which $C_i(x_j)$ was greater than a prespecified threshold t (set at the classifier boundary).

$$C_i(k) = \frac{\sum_j C_i(x_j) > t}{N_k} \tag{2}$$

Finally, we tested aggregation at the prediction level, in which we trained the gender predictor using all the semantic model vectors from all of the images in the user's feed, instead of using a single, aggregated vector for a user. We then pooled the prediction scores from the gender classifier on the images of a test user to determine his/her gender. As shown in the results in Table 1, this strategy proved to be the most effective.

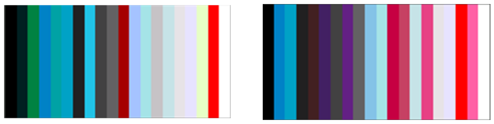
2.3. Additional Visual Information

Besides the profile picture and the images posted in the feed, a Twitter user profile contains other forms of visual information: specifically the background image, header image and profile color patterns. We therefore analyzed such content as well, and tested it in the gender classification context.

The header and background images fill the homepage of a user. Typically they are thematic pictures not containing people, and a large portion of users do not personalize them but use the default Twitter themes. In fact, in the dataset we analyzed, we found that roughly half of the users employed the default option for either the Background or the Header image. Therefore those visual clues provide weaker information with respect to other streams. We employed the Semantic Model Vector with 717 visual classifiers as the representation for both images.

Following the approach by Alowibd et al. [11], we also collected the profile color information for the following Twitter account details: Background, Text, Line, Sidebar Fill, Sidebar Fill and Sidebar Border. The information was collected using an open source service provider[5]. Each color information was encoded using color quantizations in RGB space using 8 or 9 bins per channel (resulting in codebooks of 512 and 729 elements, respectively) or directly employing

[5]http://www.twitteraccountsdetails.com/



(a) (b)

Fig. 4. Twenty most used colors by (a) male and (b) female users in Twitter profiles. Colors were quantized in 729 bins and ordered from left to right based on luminance value.

the raw color values. The gender prediction model was built on top of such representations individually and in combination using standard SVMs with RBF kernels. In the analyzed dataset, 24.21% (11.9% male and 12.31% female) of the users employed the default color options. Looking at the distributions of the 20 most used colors by male and female users in Figure 4, we notice a higher use of red, pink and brown shades in female users, whereas males seem to prefer a palette oriented to blue, green and grey.

3. TEXTUAL INFORMATION

In order to provide a comparison with the state of the art on gender prediction in social media, we also extracted and employed textual features. Note however that the purpose of this work is not to claim that visual analytics performs better gender prediction than traditional textual ones, but that it provides a solid and complementary cue that should be used in combination with existing techniques.

We used two sources of textual information, following the procedure adopted by Liu and Ruths [5], in order to try to reproduce as closely as possible the performance of their approach on the dataset they introduced and that we use in our experiments.

Tweets. We analyzed 200 tweets from each user, and learned a linear SVM on top of extracted n-grams from the text. We used the Libtext library [16] for all our processing.

First Name Analysis. We collected the first name information from the 1990 census [6] and associated each detected first name from the given profiles to its frequency within the male and/or female population.

4. MULTIMODAL INFORMATION FUSION

We tested traditional early and late fusion strategies to combine both textual and visual information.

In **early fusion**, we simply concatenated feature vectors obtained from different sources.

For **late fusion**, we tried simple pooling strategies to combine separately trained classifiers, as well training a nonlinear SVM on top of the concatenation of the prediction scores from the individual classifiers.

Since the information provided by first name analysis and profile picture face-based analytics is not encoded in a feature vector, but provides an immediate gender prediction, we also tried a **filtered fusion** approach. In this framework, the final gender prediction decision is taken immediately and without considering the other sources of information if

1) a first name matches exactly a name that associated only with either the male or female gender, or

2) the Face++ detector found only one single face and its gender prediction score is above 90%

As shown in the results reported in Tables 2 and 3, the filtered fusion strategy proved to be the most effective.

5. EXPERIMENTS

5.1. Experimental Setup

We used the dataset introduced by Liu and Ruths [5] [7], which contains 10K Twitter users and their gender information. Following their protocol, we performed 10 random splits, each containing with a test set of 800 users (400 male and 400 female), while the remaining users were used for training. Gender prediction performance was evaluated as mean accuracy over the 10 splits, with 50% representing random prediction.

All gender classifiers on top of each information vector (visual, textual, or mixed) were trained using SVMs with RBF kernel, with kernel parameters estimated via grid search. The only exception was the n-gram based textual one for which, given the extremely high dimensionality of the feature vector, we used the linear SVM classifier built in LibShortText.

5.2. Results and Discussion

From the results reported in Tables 1, 2 and 3 we can draw the following conclusions.

For the semantic scores obtained by applying visual classifiers on all the images from the whole feed of a user, learning a gender predictor using the individual images model vectors and then performing average pooling over all the image prediction scores for the test user provides better performance than aggregating the visual models scores across the images each user, and then training a gender classifier on user instances. It seems that using more categories/visual models increases the final gender classification performance, with Deep models providing the best results.

Analysis of text in tweets alone proves better than any other individual approach by a large margin. However, the performance gap between the fusion of visual information and the fusion of textual information is much smaller.

Textual and visual information are complementary, and their fusion boosts prediction accuracy.

Late filtered fusion provides the best performance, achieving 88% mean accuracy on this dataset, thus resulting in the

[6] http://www.census.gov/genealogy/www/data/1990surnames/names files.html

[7] http://www.networkdynamics.org/static/datasets/LiuRuthsMicrotext.zip

Method	Accuracy
Max Pooling	67.53
Avg Pooling	69.43
Avg Top-Quarter Pooling	69.56
Threshold-count	70.82
Avg Prediction Pooling	**71.38**

Table 1. Mean accuracy over ten fold gender prediction experiments using different aggregation methods over the images in the visual feed based on the SVM717 representation.

Method	Accuracy
Background SMV717	60.11
Header SMV717	64.41
color	66.18
Visual Feed SMV51	66.67
Visual Feed SMV717	71.38
Visual Feed SMVDeep1000	**75.40**
Profile SMV25	69.11
Profile Face++	74.90
First Name	71.22
LibText 200 Tweets	**83.37**

Table 2. Mean accuracy over ten fold gender prediction experiments using different visual and textual sources.

state of the art for such dataset. It should be noted that the results reported by Liu and Ruths [5] were obtained on different splits of the data. We expect that given their reported higher performance of textual fusion, our combination with visual information could further improve performance on the splits they employed in their experiments.

Finally in Figure 5 we report a qualitative analysis of the most discriminative visual classes for gender, selected by weight magnitude of a linear SVM trained on top of the SMV51 vectors.

6. CONCLUSIONS AND FUTURE WORK

We showed that the semantic content of the pictures posted by users in social media can be used to predict their gender. We used a set of independently trained visual classifiers, and showed through extensive experiments on a set of 10K Twitter users that such visual information can provide a strong gender predictor cue (75.6% accuracy), which proved to be complementary to traditional textual analytics (88% accuracy).

In the future, we plan to extend the use of visual information to estimate other user attributes such as age and political affiliation.

Method	Accuracy
Visual Feed Early Fusion	75.58
Visual Feed Late (avg) Fusion	74.34
Visual Feed Late (SVM) Fusion	**75.6**
Profile Late (avg) Fusion	77.85
Profile Late (SVM) Fusion	78.63
Profile Filtered Fusion	**79.05**
All Visual Late(SVM) Fusion	80.08
All Visual Late(SVM) Filtered Fusion	**83.36**
Textual Early Fusion	84.08
Textual Feed Late (avg) Fusion	84.53
Textual Late (SVM) Fusion	84.67
Textual Filtered Fusion	**85.72**
Visual+Text Early Fusion	84.07
Visual+Text Late (SVM) Fusion	85.97
Visual+Text Late (SVM) Filtered Fusion	**88.01**
Liu and Ruths [5]	87.1

Table 3. Mean accuracy over ten fold gender prediction experiments using different fusion strategies. Note that random guess produces 50% accuracy, and Liu and Ruths [5] results were obtained on different splits of the data.

7. REFERENCES

[1] John D. Burger, John Henderson, George Kim, and Guido Zarrella, "Discriminating gender on twitter," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2011, EMNLP '11, pp. 1301–1309.

[2] Athanasios Kokkos and Theodoros Tzouramanis, "A robust gender inference model for online social networks and its application to linkedin and twitter," *First Monday*, vol. 19, no. 9, 2014.

[3] Shane Bergsma and Benjamin Van Durme, "Using conceptual class attributes to characterize social media users.," in *ACL (1)*. 2013, pp. 710–720, The Association for Computer Linguistics.

[4] Shuo Chang, Vikas Kumar, Eric Gilbert, and Loren Terveen, "Specialization, homophily, and gender in a social curation site: Findings from pinterest.," in *CSCW*, 2014.

[5] Wendy Liu and Derek Ruths, "What's in a name? using first names as features for gender inference in twitter.," in *AAAI Spring Symposium: Analyzing Microtext*. 2013, vol. SS-13-01 of *AAAI Technical Report*, AAAI.

[6] D. Nguyen, D. Trieschnigg, A.S. Doğruöz, R. Gravel, M. Theune, T. Meder, and F. de Jong, "Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment," in *Proceedings of COLING*, 2014.

[7] Morgane Ciot, Morgan Sonderegger, and Derek Ruths, "Gender inference of Twitter users in non-English contexts," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Washington, USA, October 2013, pp. 1136–1145, Association for Computational Linguistics.

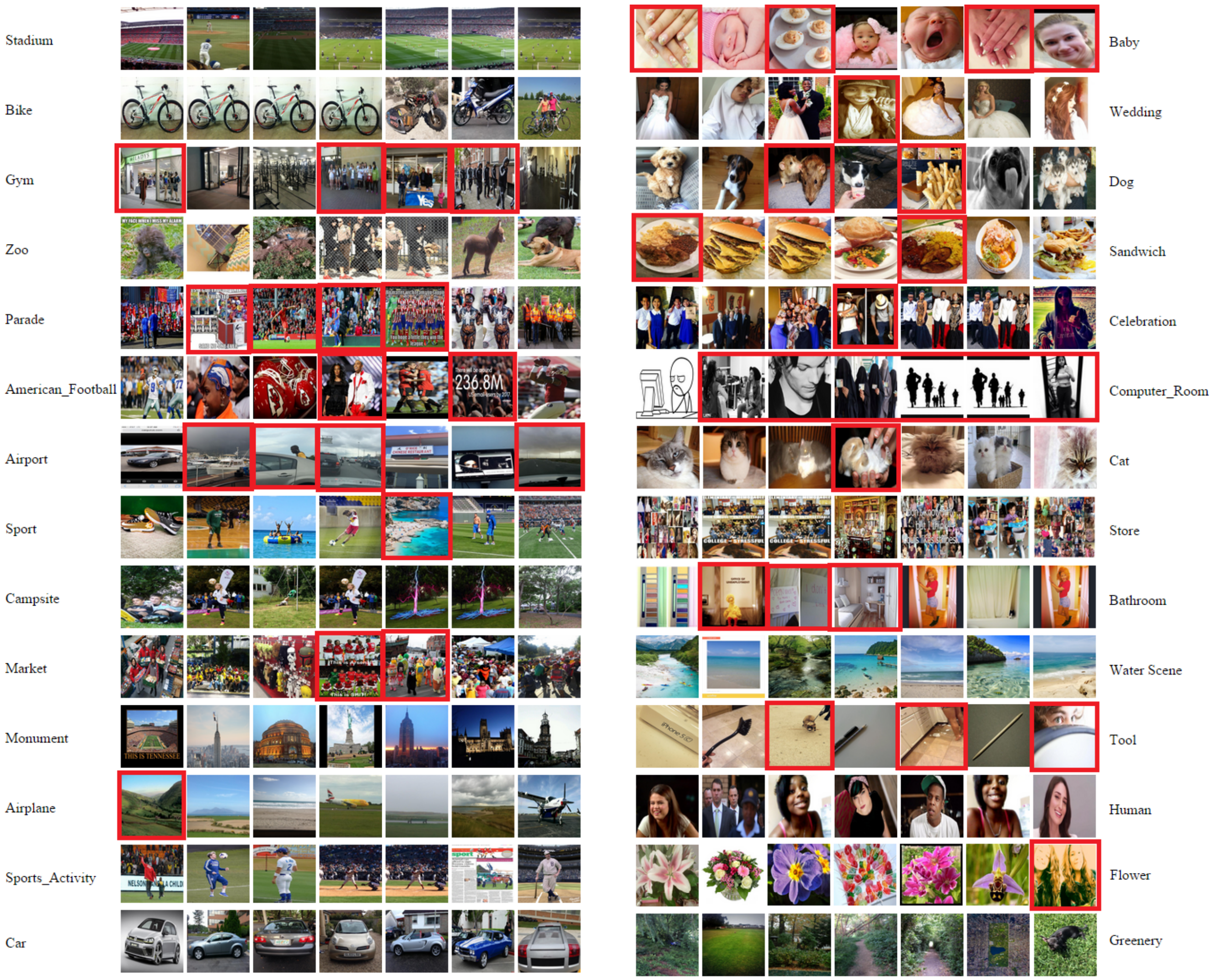


Fig. 5. Top seven retrieved images for the most discriminative categories for male (left) and female (right). Images marked in red represent classifiers errors.

[8] J. Ito, T. Hoshide, H. Toda, T. Uchiyama, and K. Nishida, "What is he/she like?: Estimating twitter user attributes from contents and social neighbors," in *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, 2013, pp. 1448–1450.

[9] Faiyaz Al Zamal, Wendy Liu, and Derek Ruths, "Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors.," in *ICWSM*. 2012, The AAAI Press.

[10] Puneet Singh Ludu, "Inferring gender of a twitter user using celebrities it follows," *CoRR*, vol. abs/1405.6667, 2014.

[11] Ugo A. Buy Jalal S. Alowibdi and Philip S. Yu, "Language independent gender classification on twitter," in *The 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM'13*, 2013.

[12] Marco Pennacchiotti and Ana-Maria Popescu, "A machine learning approach to twitter user classification.," in *ICWSM*. 2011, The AAAI Press.

[13] Xiaojun Ma, Yukihiro Tsuboshita, and Noriji Kato, "Gender estimation for sns user profiling using automatic image annotation.," in *ICME Workshop on Cross-media Analysis from Social Multimedia (CASM)*, 2014, pp. 1–6.

[14] Haoqiang Fan, Mu Yang, Zhimin Cao, Yuning Jiang, and Qi Yin, "Learning compact face representation: Packing a face into an int32," in *ACM Multimedia*, 2014.

[15] Michele Merler, Bert Huang, Lexing Xie, Gang Hua, and Apostol Natsev, "Semantic model vectors for complex video event recognition," *Multimedia, IEEE Transactions on*, vol. 14, no. 1, pp. 88 –101, feb. 2012.

[16] H.-F. Yu, C.-H. Ho, Y.-C. Juan, and C.-J. Lin., "Libshorttext: a library for short-text classification and analysis.," in *Technical report*, 2013.