A Weighted Combination of Text and Image Classifiers for User Gender Inference

Tomoki Taniguchi, Shigeyuki Sakaki, Ryosuke Shigenaka,
Yukihiro Tsuboshita and Tomoko Ohkuma
Fuji Xerox Co., Ltd. , 6-1, Minatomirai, Nishiku, Yokohama-shi, Kanagawa, Japan
{taniguchi.tomoki,sakaki.shigeyuki,shigenaka.ryosuke,
yukihiro.tsuboshita,ohkuma.tomoko}@fujixerox.co.jp

Abstract

Demographic attribute inference of social networking service (SNS) users is a valuable application for marketing and for targeting advertisements. Several studies have examined Twitter-user gender inference in natural language processing, image recognition, and other research domains. Reportedly, a combined approach using text data and image data outperforms an individual data approach. This paper presents a proposal of a novel hybrid approach. A salient benefit of our system is that features provided from a text classifier and from an image classifier are combined appropriately to infer male or female gender using logistic regression. The experimentally obtained results demonstrate that our approach markedly improves an existing combination-based method.

1 Introduction

Concomitantly with rapid growth in SNS, consumers increasingly use SNS to exchange and share their opinions related to products, services, politics, and other matters. Many companies are motivated to use SNS data for marketing or advertisement to satisfy needs for improvements of their products or services in real time with low cost. However, in many cases, SNS user information such as gender, age or residence is not openly available, although such information is extremely important for marketing. To meet that objective, several studies have been conducted to infer demographic information of anonymous users using text or image data posted on Twitter, and community membership (Rao and Yarowsky, 2010; Ikeda et al., 2013; Ma et al., 2014; Sakaki et al., 2014). Sakaki et al. (2014) demonstrated that a hybrid-based method outperformed other approaches using individual sources. However we observed an important issue: each probability score output from the image classifiers and the text classifier was simply summed, although the degree of their respective contributions to the inference is presumably different.

As described herein, considering that issue, we propose a novel method with a hybrid approach using logistic regression. In addition, from examination of experimentally obtained results, we show which image contents contribute strongly to the inference of a Twitter user being male or female.

2 Related Work

Earlier studies investigated demographic attribute inference for SNS users based on machine learning. Text and images posted on SNS, membership in virtual communities, and combined information have been used as training data.

Burger et al. (2011) and Liu et al. (2012) applied text to infer user demographic attributes. Burger et al. (2011) realized a classifier that discerns SNS user gender. Liu et al. (2012) estimated the gender makeup of commuting populations using text.

Ma et al. (2014) and Ulges et al. (2012) used images and videos posted on SNS. Ma et al. (2014) defined 30 sub-categories, which were combinations of 10 image contents and 3 gender attributes (male, female, and unknown), and described a system that inferred a user's gender by classifying posted images into sub-categories. Ulges et al. (2012) detected TV viewers' gender and age via content-based concept detection.

Ikeda et al. (2013) and Sakaki et al. (2014) used methods that incorporate information. Ikeda et al. (2013) proposed a hybrid-based method using both text and community membership. Sakaki et al. (2014) proposed a hybrid-based method using a combination of text and images, which builds a meta-classifier using the probability score out-

Proceedings of the 2015 Workshop on Vision and Language (VL'15), pages 87–93,
Lisbon, Portugal, 18 September 2015.

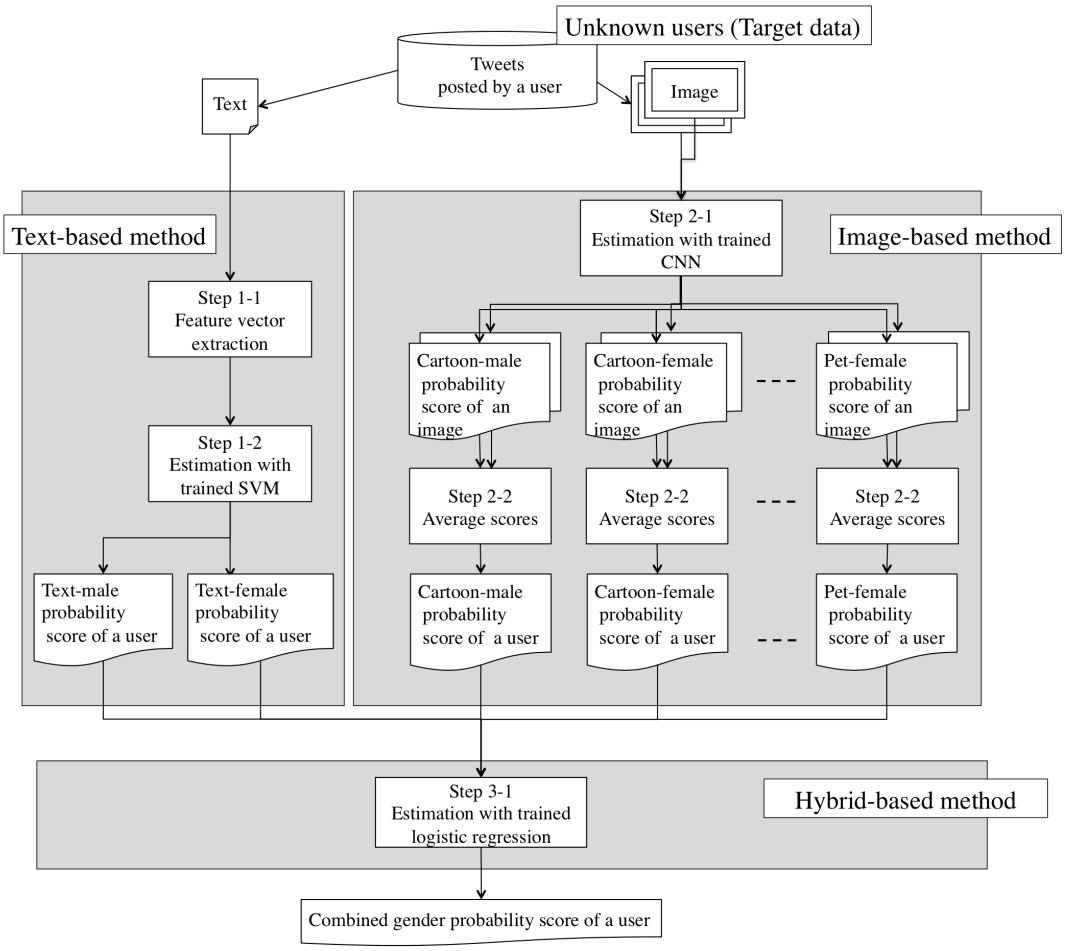


Figure 1: Overview of our proposed method.

put from text and image classifiers as input. This study demonstrated that a combination of text and images boosts the accuracy of a single source.

Since Krizhevsky et al. (2012) won first prize overwhelmingly at ILSVRC-2012, Convolutional Neural Networks (CNN) has gained great attention in the research field of image classifications. With the rise of efficient GPU computing, CNN has been used in practical applications. A few reports have described applications of CNN, which deals with inference of user attributes. Shigenaka et al. (2015) applied CNN for gender inference, demonstrating that CNN performs much better than a classifier based on SVM.

3 Proposed Method

Figure 1 presents an overview of our proposed method. Our method classifies some attributes (here, genders) of people who posted text and images. Our method includes three component methods that are: text-based, image-based, and hybrid-based.

3.1 Text-Based Method

The text-based method receives text as input and outputs the male and female probability scores. We used SVM to classify genders. To retrieve probability scores, we used logistic regression. The logistic function converts a distance from a hyper plane to probability scores of 0.0 - 1.0. As shown in Equation 1, the sum of the male and female probability scores is 1. The text-based method procedure is the following.

Step 1-1 Tokenization is done using Kuromoji (`http://www.atilika.org`), a Japanese morphological analyzer. Thereby, the unigram is obtained. Then, the bag-of-words feature is extracted from the unigram.

Step 1-2 The SVM receives the bag-of-words fea-

ture as input. The male probability score is obtained using SVM. Then, the female probability score is calculated using Equation 1.

$$score_{male} + score_{female} = 1 \quad (1)$$

3.2 Image-Based Method

The image-based method classifies the contents of posted images and estimates the gender of people who posted them. The image-based method receives an image as an input and outputs the probability score of each sub-category. Sub-categories are defined as combinations of image contents and user attributes: in this study, genders. Details of the sub-categories are presented later in section 4.1. We used a CNN model comprising 16 layers (Simonyan and Zisserman, 2014), which is pre-trained using the ILSVRC-2012 corpus. Neurons of the output layer of the pre-trained model are replaced with the same numbers of neurons as sub-categories. Weights of the new output layer are initialized to random values. Then, the weight of the pre-trained model is fine-tuned with the training dataset using backpropagation of error derivatives. Details of the dataset are presented later in section 4.1.

The image-based method procedures are the following.

Step 2-1 Probability score of images for sub-categories is obtained using the CNN model.

Step 2-2 The score of each user is obtained by averaging the probability score of images that the user posts since many users posted more than one image.

3.3 Hybrid-Based Method

The hybrid-based method classifies scores related to text-based and image-based method output and estimates the gender of people who posted text and images. We used logistic regression. In step 3-1, the male probability score is obtained using logistic regression. Then, the female probability score is calculated using Equation 1 in the same manner as that presented in step 1-2.

The training process for logistic regression involves two stages. In the first stage, the text-based and image-based method are trained to obtain training data for the hybrid-based method. In the second stage, the hybrid-based method is trained using them.

4 Experiment

We conducted a gender classification on Twitter.

4.1 Experimental Data

Experimental data are of two levels: a tweet level and an image level. We prepared a huge number of annotation data as a training corpus using Yahoo Crowd Sourcing (`http://crowdsourcing.yahoo.co.jp/Yahoo`).

Tweet level annotation: Tweet level annotation process refers to rules proposed by Sakaki et al. (2014), who defined the tweet level labels as male and female. Workers annotated the labels using many sources of potentially discriminative meta-data, including user preferences, icons, text, and images.

Image level annotation: The image level annotation process refers to rules proposed by Ma et al. (2014), who defined image labels as the combination of the gender of users who had posted images and the contents that the images are likely to express. The image labels include two parts. The first is a gender category: female, male, and unknown. The former two are used to label images, for which people can infer the uploader gender. For images of which the uploader gender is unrecognizable, we use unknown. The second part defined in the image label is the category that expresses the classification of contents included in images. We designate the combination of these categories as sub-category. Table 1 shows typical contents of the sub-category.

Finally, we obtained 6000 tweet level annotations and 8162 image level annotations. As shown in Figure 2, the tweet level annotation data were split up into three subsets. Subset A was used for training the text-based method. Subset B was used for training the hybrid-based method. Subset C was used for evaluation. Image level annotation data were used for the training-image-based method.

4.2 Experimental Setup

LIBSVM (Chang and Lin, 2001) was used as the implementation of SVM. The linear kernel was selected. Then LIBLINEAR (Rong-En et al., 2008) was used as the implementation of logistic regression. Cost parameter C was set to 1.0.

(a) baseball stadium

(b) barbecue

(c) shaved ice

		Gender category		
		female	male	unknown
Contents category	cartoon	Romance cartoon	Hero cartoon	Unisex cartoon
	famous people	Famous male idol	Famous female idol	Comedian
	food contents	Shaved ice	Barbecue	Sandwich
	consumer goods	Jewelry	Electrical appliances	Cellular phone
	memo	Colorful memo	Black and white memo	Short memo
	outdoor	Amusement park	Baseball stadium	Landscape
	person	Girl,woman,baby	Boy,man	Crowd of people
	pet	Penguin,small dog	Frog,tiger	Cat
	screenshot	Pastel color screen	TV game screen	Weather news
	others	Beauty advertisement	Transportation	Black screen

Table 1: Sub-category composed with combinations of the gender and contents category. This table shows typical contents of the sub-category obtained using image-level annotation. For example, the combination of male and outdoor includes a baseball stadium image (a), the combination of male and food contents includes a barbecue image (b), and the combination of female and food contents includes the image of a shaved ice (c).

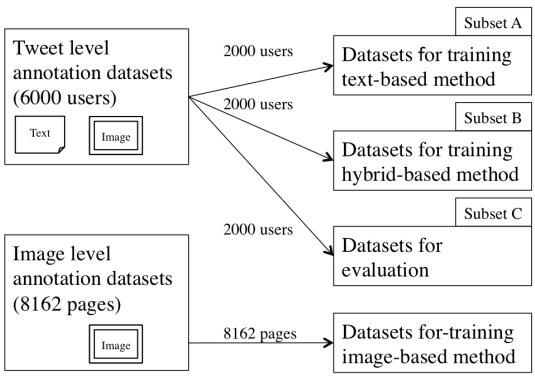


Figure 2: Datasets for training and evaluation.

As comparative methods, we selected the method presented by Sakaki et al. (2014) using the combination approach of text and image data and also the selected text-based and image-based method using the approach of a single source. In the experiment, classifiers of Sakaki et al. (2014) were replaced with the proposed classifiers to compare the performances of hybrid methods. The alpha value necessary for the method of Sakaki et al. (2014) to combine probability scores was set to 0.74 based on preliminary experiments.

5 Experimental Results

Table 2 shows the precision, recall, F-measure, and accuracy. The accuracy of our proposed method achieved 80.25 [%], which is 2.95 pt higher than that of the text-based method, 8.25 pt higher than that of the image-based method, and 1.35 pt higher than that of the method described by Sakaki et al. (2014). Especially, the female F-measure associated with our proposed method achieved 77.07 [%], which is 5.03 pt higher than that of the text based method, 13.69 pt higher than that of the image based method, and 2.0 pt higher than that of the method described by Sakaki et al. (2014).

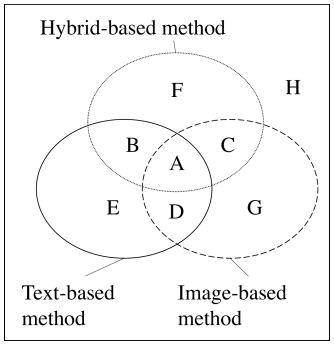
We conducted a binomial test to assess our proposed method and the method described by Sakaki et al. (2014). Results confirmed that the p value is 0.0031, which indicates that the results obtained using our method are significantly better than those obtained using the existing combination-based method.

	Male			Female			
	Precision	Recall	F-measure	Precision	Recall	F-measure	Accuracy
Text-based method	76.20	86.12	80.88	79.16	66.10	72.04	77.30
Image-based method	70.41	85.97	77.41	75.58	54.58	63.38	72.00
Sakaki et al. (2014)	77.66	86.99	82.06	80.69	68.47	75.07	78.90
Proposed method	80.92	84.48	82.66	79.36	74.91	77.07	80.25

Table 2: Experimental results.

6 Discussion

This section presents a discussion of the effectiveness of the combination of the text-based and image-based methods. Then the discussion addresses the difference between the model proposed by Sakaki et al. (2014) and our proposed method. Finally, the applications of logistic regression weights of the combined sources are discussed.



Set	T	I	P
A	⋆	⋆	⋆
B	⋆		⋆
C		⋆	⋆
D	⋆	⋆	
E	⋆		
F			⋆
G		⋆	
H			

Figure 3: Venn diagram: T denotes text-based method. I denotes image-based method. P denotes hybrid-based method. "⋆" denotes the set of users whose genders were inferred correctly using each method.

	Number of users		
Set	Conventional	Proposed	Difference
A	1105	1097	-8 ⇓
B	408	412	+4 ⇑
C	62	78	+16 ⇑
D	0	8	+8 ⇑
E	33	29	-4 ⇓
F	1	18	+17 ⇑
G	180	164	-16 ⇓
H	211	194	-17 ⇓

Table 3: Number of users included in each set. Conventional approach denotes the method of Sakaki et al. (2014).

6.1 Effectiveness of the Combination Approach

Figure 3 portrays the relation between the text-, image-, and hybrid-based methods. Each circle of the Venn diagram represents a set of users whose gender was inferred correctly using a method. The union of A, B, D, and E represents users whose gender was inferred correctly using the text-based method. B represents users whose gender was inferred correctly by the text-based and the hybrid-based method, but was misjudged using the image-based method.

Table 3 presents the number of users each set contains. We would like to examine C, D, E, and F specifically to assess the difference in the performance between the text-based and the hybrid-based method. C and F include users whose respective genders were inferred correctly using the hybrid-based method, but misjudged using the text-based method. D and E include users whose respective genders were inferred correctly using the text-based method, but misjudged using the hybrid-based method. Here we discuss the results obtained using the proposed method, which are shown at 3rd column. Regarding C, D, E, and F, C includes the maximum number of users, 78 (3.9 [%]), whose respective genders were inferred correctly using the proposed method. For C, a user's gender was inferred correctly using the image-based method. Therefore, it is apparent that our proposed method increased the number of correct answers considerably by taking in the correct region of the image-based method. It is particularly interesting that although the text-based and image-based methods both misjudged users (18 users : 0.9 [%]) in F, the proposed method can infer their genders correctly. However, D and E include only 37 users (8 + 29 users : 1.85 [%]) whose gender was misjudged using our proposed method. Therefore, our proposed method increased the number of correct inferences

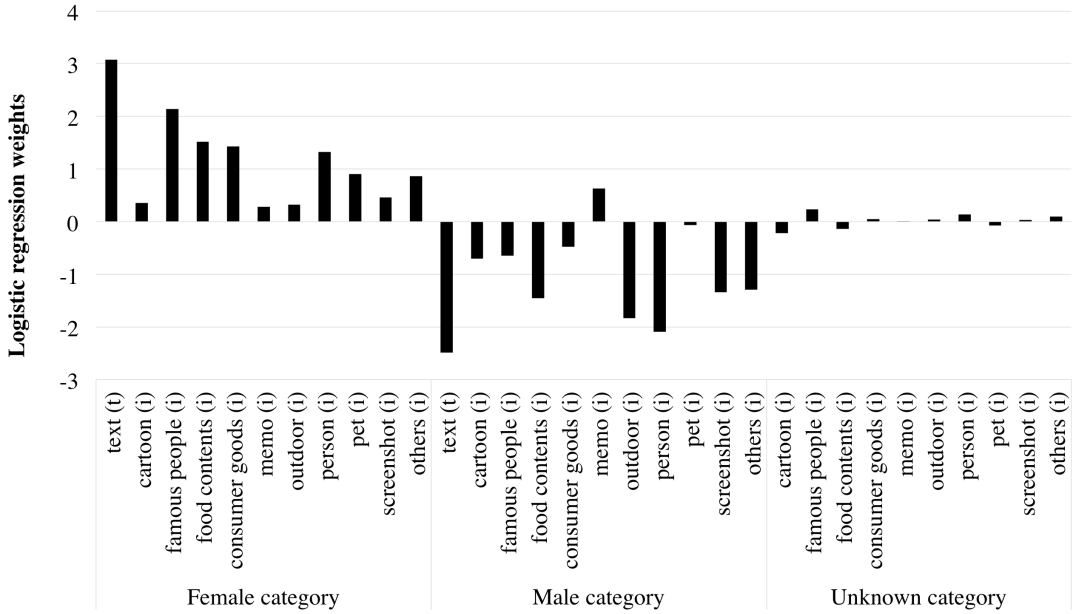


Figure 4: Logistic regression weights. Character (t) denotes weights with respect to the text-based method; (i) denotes weights with respect to the image-based method.

Rank	Male	Absolute weight	Female	Absolute weight
1_{st}	person	2.08	famous people	2.13
2_{nd}	outdoor	1.82	food contents	1.51
3_{rd}	food contents	1.45	consumer goods	1.42

Table 4: Top three absolute weights of the image content categories.

(59 users : 2.95 [%]) beyond the level of the individual text-based method.

6.2 Comparison of the Conventional Approach

The difference between the proposed method and that proposed by Sakaki et al. (2014) can be discussed with reference to Figure 3 and Table 3. Except for H, which includes users whose gender was misjudged by all methods, the difference between our method and the method presented by Sakaki et al. (2014) in F was the largest (+17 users). Actually, F includes users whose gender was newly inferred correctly by the hybrid-based method but whose gender was misjudged using individual methods. Therefore, our method more correctly infers a new user's gender by combining sources of text and images than the method presented by Sakaki et al. (2014). We assume that our method handles the combination appropriately to infer male or female gender using logistic regression.

The difference between our method and that presented by Sakaki et al. (2014) in C was the second largest (+16 users). Actually, C includes users whose gender was inferred correctly using the hybrid-based and the image-based method, but misjudged using the text-based method. Therefore, we assumed that our proposed method handled the image source more appropriately than the method presented by Sakaki et al. (2014).

6.3 Logistic Regression Weights

Figure 4 shows the logistic regression weights. From this figure, we observed that the weights for female users were all positive, the weights for male users were almost all negative. The weights for unknown were nearly zero, which indicates that the probability scores of text-based and image-based methods are not competing.

Presumably, the logistic regression weights obtained by training indicate the rate of the contribution to the inference. Table 4 presents the top three image content categories according to their absolute weights. The table shows that person, outdoor, and food contents are clues to male gender,

but famous people, food contents, and consumer goods imply female gender.

Consequently, through analysis of the logistic regression weights, we confirmed that the rates of image contents' contributions to inference mutually differed. We therefore conclude that the rate of each image content's contribution to the inference is expected to be different for different genders. Our proposed method performs significantly better than the existing combination-based method.

7 Conclusion

This paper presented a proposal for a novel hybrid approach. The salient benefit of our system is that features provided from a text classifier and from an image classifier are combined appropriately to detect male or female gender using logistic regression. Experimental results show that our approach achieved accuracy of 80.25 [%], which was 1.35 pt higher than the conventional combination approach. In addition, through analysis of logistic regression weights, we confirmed that the rate of each image content's contribution to the inference should be different for different genders. Person, outdoor, and food contents are clues to male gender, but famous people, food contents, and consumer goods imply female gender. We therefore conclude that our proposed method using weighted combination of text and image classifiers performs markedly better than existing combination method.

Our approach is applicable to other attributes that might be inferred for SNS users, such as age, career, and residence, which were investigated by Ikeda et al. (2013) and by Rao and Yarowsky (2010). Because it is presumed that posted image contents clearly reflect SNS user hobbies and lifestyles, our approach is suitable for inferring those attributes as well. As a subject for future work, we intend to apply our approach to the inference of various SNS user attributes.

References

Adrian Ulges, Markus Koch and Damian Borth. 2012. Linking visual concept detection with viewer demographics. In *Proceedings of the ACM International Conference on Multimedia Retrieval*.

Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1106-1114.

Chin-Chung Chang and Chin-Jen Lin. 2001. LIBSVM: a library for support vector machines. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`.

Delip Rao and David Yarowsky. 2010. Detecting latent user properties in social media. In *Proceedings of the Neural Information Processing Systems Workshop on Machine Learning for Social Networks*.

John D. Burger, John Henderson, George Kim and Guido Zarrella. 2011. Discriminating gender on Twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1301-1309.

Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale visual recognition. Software available at `http://www.robots.ox.ac.uk/~vgg/research/very_deep/`.

Kazushi Ikeda, Gen Hattori, Chihiro Ono, Hideki Asoh and Teruo Higashino. 2013. Twitter user profiling based on text and community mining for market analysis. In *Knowledge Based Systems*, pages 35-47.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang and Chin-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. Software available at `http://www.csie.ntu.edu.tw/~cjlin/liblinear/`.

Ryosuke Shigenaka, Yukihiro Tsuboshita and Noriji Kato. Image-based user gender inference using deep learning on SNS. In *Proceedings of the Meeting on Image Recognition and Understanding*, in Japanese.

Shigeyuki Sakaki, Yasuhide Miura, Xiaojun Ma, Keigo Hattori and Tomoko Ohkuma. 2014. Twitter user gender inference using combined analysis of text and image processing. In *Proceedings of the workshop of the International Conference on Computational Linguistics*, page 54-61.

Wendy Liu, Faiyaz Al Zamal and Derek Ruths. 2012. Using social media to infer gender composition of commuter populations. In *Proceedings of the International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media*.

Xiaojun Ma, Yukihiro Tsuboshita and Noriji Kato. 2014. Gender estimation for SNS user profiling automatic image annotation. In *Proceedings of the International Workshop on Cross-media Analysis for Social Multimedia*.