



# Finding Baby Mothers on Twitter

Yihong Zhang<sup>1</sup>(✉), Adam Jatowt<sup>1</sup>, and Yukiko Kawai<sup>2</sup>

<sup>1</sup> Kyoto University, Kyoto, Japan

yh Zhang7@gmail.com, adam@dl.kuis.kyoto-u.ac.jp

<sup>2</sup> Kyoto Sangyo University, Kyoto, Japan

kawai@cc.kyoto-su.ac.jp

**Abstract.** In this paper, we study the task of detecting mothers of babies on Twitter. This could be beneficial for baby mother users to find friends, and for companies, organizations or experts to deliver accurately targeted information. Prior works have proposed supervised classification methods to detect generic latent attributes of Twitter users such as age, gender, and political orientation. However, methods and features for classifying generic attributes do not perform well for more specific attributes, such as whether a user is a mother of a young baby. We design feature sets based on followed accounts and profile pictures, which are largely overlooked in existing work. Comparing to three established feature sets, the experimental evaluation shows that our specifically-designed feature sets considerably improve classification accuracy.

**Keywords:** Twitter · Social media analysis · User discovery

## 1 Introduction

On a social media websites such as Twitter, there is a wide range of information associated with users, including profile description, profile picture, home location, past tweets, and follower-following connections<sup>1</sup>. However, such platform often does not explicitly provide user attributes such as gender, age, and various other user characteristics. Hence a number of existing works have proposed methods to predict these attributes [1, 6, 8]. Recently, detecting special minor groups of users has also attracted considerable attention. For example, dedicated approaches have been proposed to detect users who suffer depression or users who are in various phases of drug addiction [2–4]. With regard to parenting, Morris investigates usage of social media such as Twitter and Facebook by mothers of young children [5], which provides findings related to our work.

In this paper, we present a solution for automatically detecting *mothers who have baby children* on Twitter. Similar to Morris' study, we define a baby as a child that is less than three years old [5]. As Morris's study shows, mothers with baby children tend to often actively seek information and advice from baby caring experts, facility providers, and other mothers in the same situation.

<sup>1</sup> <https://dev.twitter.com/overview/api/users>.

We can devise various applications in the face of such needs. For example, we can generate information on which streets are frequently used by other mothers with young babies by detecting baby mothers in social media, and this information can be then used for specialized route recommendation. The pilot interview we conducted with baby mothers within our research group revealed that, baby mothers are willing to be provided with such applications. In general, our method for automatic detecting mothers having babies will be beneficial for mothers who wish to contact other mothers in similar situation, and for experts, as well as organizations who aim to provide targeted and on-time information.

In the existing works on Twitter users classification, the most often used information are users' past tweets [6, 8, 10]. However, as Morris has found in her study, baby mothers who use Twitter often do not share their baby information in tweets, except for profile pictures [5]. Based on our investigation, a Twitter user who is a mother to a baby would nevertheless seek baby-related information by following expert and consultant accounts. Therefore, instead of using information contained in users' tweets as in the prior works, we focus on the accounts a user follows. We devise a set of new features based on followed accounts and some other information. We manually select some baby mothers and non-baby mothers from real Twitter users for training and testing our prediction model. According to experimental results, the classification accuracy using our proposed features is higher than using the established feature sets.

## 2 Related Work

In recent years, discovering latent Twitter user attributes has become an important research topic. For instance, Rao *et al.* [8] propose a machine learning approach to discover gender, age, place of origin, and political orientation based on features extracted from users' past tweets. Pennacchiotti and Popescu [6] propose another machine learning approach to discover political affiliation, ethnicity, and to find Starbucks fans among Twitter users based on profile, tweet behavior, and tweet contents. Sharma *et al.* [9] introduce a method to infer keywords associated with the biographical information and expertise of a user. However, it is unclear whether these generic methods can be effective in our task of discovering baby mothers. We compare this generic method with our own approach in terms of classification accuracy.

Recently, detecting special groups of social media users who are in need also attracts researches. MacLean *et al.* study users with drug addition problems in a online forum [4]. Their focus is on detecting users in different phases of drug addition, including using, withdrawing, and recovering. De Choudhury *et al.* study Twitter users who suffer depression [3]. They propose a detection method based on several signals including tweet posting activities, social network graph, lexicon-based emotion analysis, linguistic style, and special words used by patients. Among several classifiers, they found that Support Vector Machine (SVM) provides best prediction results. The authors make a similar study of mothers who recently gave birth [2]. The features they used include engagement, network, emotion, and linguistic style. Their prediction mainly focuses

on detecting changes occurring between the time of prepartum and postpartum, and it is unclear if it is effective for detecting baby mothers among different types of users. We also compare that method with our own approach. Nevertheless, to the best of our knowledge, our work is the first proposal of categorizing SNS users according to whether they are mothers to a baby.

### 3 Features for Finding Baby Mothers

We follow the approach for detecting latent attributes of social media users using machine learning, which include devising features, labeling examples, and testing models. Our main contribution is that we not only propose a novel task but also introduce specific novel features effective for detecting mothers with babies. In this section, we discuss the features we consider in our framework, which include existing and novel features. The data preparation and learning models will be discussed in the next section.

#### 3.1 Existing Features

**Bag-of-Words on Tweets (BOW).** The simple presentation of users' text as term frequency has been a baseline method in several user classification works, including author profiling task in PAN competitions 2017 [7]. Following [7], we use BOW of 1,000 most frequent terms to represent a user's tweets.

**Socio-Linguistic.** We study a feature set for generic attribute detection. Proposed in [8], this feature set is used with SVM for detecting gender, age, regional origin and political orientation. The entire feature set is listed in Table 1. Most of these features are based on lexicons. Some lexicons are pre-defined, others are generated from the data (e.g., possessive bigrams). The feature value indicates the count of occurrences of the lexicon words in the user's tweets.

**Postpartum.** The second feature set we use is specialized for detecting behavior of new mothers. Proposed in [2], the feature set includes posting behavior, ego-network, and linguistic style. The complete feature set is listed in Table 2. Designed to distinguish standard and extreme behavior changes for new mothers, this is the closest feature set we find to our goal of detecting baby mothers. This feature set considers tweeting activities such as retweeting, mentioning, and linking, as well as simple social network features such as the number of followers and followees. The main part of the feature set, however, is writing style analysis based on LIWC<sup>2</sup> and ANEW<sup>3</sup> lexicon.

#### 3.2 Novel Features

**Interest Keywords.** A baby mother will have a strong tendency to seek information about health advices, baby-related event news, and experiences of other mothers. We consider that it is not what the user tweets about but what she

<sup>2</sup> <http://liwc.wpengine.com/>.

<sup>3</sup> <http://csea.php.ufl.edu/media/anevmessage.html>.

**Table 1.** Socio-Linguistic features for generic user attribute detection

Feature name	Length	Description/example
SMILEYS	124	A list of emoticons compiled from the Wikipedia
OMG	1	Abbreviation for “Oh My God”
ELLIPSES	1	“...”
POSSESSIVE BIGRAMS	by data	E.g. my XXX, our XXX
REPEATED ALPHABETS	1	E.g. niceeeeeee, noooo waaaay
SELF	by data	E.g., I xxx, Im xxx
LAUGH	5	E.g. LOL, ROTFL, LMFAO, haha, hehe
SHOUT	1	Text in ALLCAPS
EXASPERATION	5	E.g. Ugh, mmmm, hmmm, ahh, grrr
AGREEMENT	3	E.g. yea, yeah, ohya
HONORIFICS	4	E.g. dude, man, bro, sir
AFFECTION	1	E.g. xoxo
EXCITEMENT	1	A string of exclamation symbols (!!!!!)
SINGLE EXCLAIM	1	A single exclamation at the end of the tweet
PUZZLED PUNCT	1	A combination of any number of ? and ! (!?!?!?)

wants to read that provides the hint on the user interest. In Twitter, a user will see most often the tweets posted from the accounts that they follow. So instead of the tweets a user posted, we are interested in the accounts a user chose to follow. Our first feature set is based on the descriptions of accounts followed by a user. Specifically, we generate *tf-idf* scores for the words used in the followed account descriptions. To use *tf-idf*, we first extract frequently used keywords from all account descriptions we have in the data, resulting in a list of  $m$  keywords. Then we concatenate descriptions of all accounts a user followed as one document. Last, we generate a *tf-idf* vector of length  $m$ . In our experiments, we set the frequency thresholds to 100, resulting in a keyword list of 1,370 words. We find that different threshold values produce similar classification results.

**Mom-Words.** We generate a list of words likely to be used by baby mothers, not necessarily baby-related. For this, we choose a popular online forum about parenting of baby children, and we collect messages posted in it. The forum we selected is *MomForum.com*, which is a dedicated sharing platform for discussing parenting and baby-related issues. We fetch 100 threads from one of the sub-forums that discusses parenting of 0–2 years old babies<sup>4</sup>. We then tokenize

<sup>4</sup> [http://www.momforum.com/forumdisplay.php/76-Babies-amp-Toddlers-\(0-2-years-old\)](http://www.momforum.com/forumdisplay.php/76-Babies-amp-Toddlers-(0-2-years-old)).

**Table 2.** Postpartum features for detecting postpartum behavior of new mothers

Feature name	Length	Description/example
Engagement	5	Average normalized number of posts per day
		Mean proportion of reply posts
		Fraction of retweets
		The proportion of links
		Fraction of question-centric posts
Ego-network	2	The number of followers
		The number of followees
Emotion	4	Positive emotion - fraction of positive emotion words in LIWC
		Negative affection - fraction of negative emotion, anger, anxiety, and sadness words in LIWC
		Mean of activation per post based on ANEW
		Mean of dominance per post based on ANEW
Linguistic style	22	Fraction of words from 22 selected LIWC linguistic styles

and generate a list of frequent keywords from these threads using a frequency threshold of 5 and removing stopwords. As the result, we obtain a lexicon of 299 words, including *child*, *feeding*, *dad*, etc. Based on this lexicon, we generate a *Mom-word* usage feature vector for each user, where the  $i$ -th element is the frequency of  $i$ -th word in the lexicon occurring in the user’s tweets.

**Pictures.** We also deploy the latest image processing techniques to extract information from user profile pictures and generate a picture feature vector. We use a free online API<sup>5</sup> that has the capability to output the age and the gender of people shown in an image with relatively high accuracy. For the particular application of discovering baby mothers, we are most interested in whether the picture shows a baby. We generate a vector that includes four binary values, each indicating whether the profile picture contains a baby ( $age \leq 3$ ), a young child ( $3 < age \leq 6$ ), an adult woman ( $20 < age \leq 50, gender = female$ ), and an adult man ( $20 < age \leq 50, gender = male$ ). Note that there can be several people of different age and or gender present in a picture.

## 4 Experimental Analysis

We conduct an empirical study by testing our feature sets on a number of real Twitter user data. We aim to find out the effectiveness of our feature sets in comparison to the established user-classification feature sets. In this section, we present the experimental setup and discuss the results.

<sup>5</sup> <https://www.angus.ai/>.

## 4.1 Dataset Preparation

We collect from Twitter a number of user profiles, and manually label them according to whether they belong to baby mothers or not. Since the proportion of baby mothers is small among all Twitter users, instead of randomly collecting user accounts on Twitter, we use some heuristics to select candidates before labeling. First we find some *attraction* Twitter account based on several handpicked Web articles<sup>6</sup>. These accounts post mother-related information, and should contain a large portion of baby mothers in their followers. Using Twitter API for searching followers<sup>7</sup>, we collect all followers from the picked accounts, including their account names, descriptions, and the numbers of followers and followees, resulting in the set of 18,536 users and their data. From these users, we first remove ones who have more than 200 followers. This considers both Twitter API limit<sup>8</sup> and the practicality of our solution, because celebrity users who have a large number of followers cannot be considered as typical users, and will distort the classification model.

We next extract a number of *very likely* candidates from remaining users whose profile description contains keyword “mother” or “mom”. Note that this filtering does not bias the classification because features used for building classification models do not consider these profile descriptions. Then from the followers of the very likely candidates, we find a number of *less likely* candidates, for which we do not use any filtering. The intuition is that the followers of a baby mother might include other baby mothers, among family and friends. Finally, we manually label positive and negative examples from these very likely and less likely candidates, by looking at all available user information, including profile description, picture, and past tweets. For very likely candidates, we have 197 positives and 59 negatives, while for less likely candidates, we have 106 positives and 583 negatives. In total, we have 303 positives and 642 negatives.

## 4.2 Learning Model and Evaluation Matrix

Rao *et al.* finds that the Support Vector Machine (SVM) with linear kernel provides the best classification results with Socio-Linguist features [8], while De Choudhury *et al.* finds that SVM with radial kernel provides the best result with their postpartum features [2]. As such, in experiment results, we show SVM with linear kernel results, except for feature sets that include postpartum features, for which we use SVM with radial kernel results. We also investigate other machine learning models including *Naive Bayes*, *Random Forest*, *Linear Discriminant Analysis*, and *Logistic Regression*, and find that Random Forest (RF) generally

<sup>6</sup> For example, *17 Funny Moms on Twitter*, <http://mashable.com/2013/05/10/funny-twitter-moms/>. Particularly, we pick four accounts from this article: @jennawrites, @laneymg, @shriekhouse, and @MarinkaNYC. The selection considers the number of followers in these accounts and Twitter API limit for collecting followers.

<sup>7</sup> <https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-followers-list>.

<sup>8</sup> Currently Twitter API allows searching 200 followers each minute.

provides the best results and performs better than SVM. Thus, we show the results for SVM and RF. We use the SVM and RF implementation in R package `e1071`<sup>9</sup> and `randomForest`<sup>10</sup>.

We measure the *precision*, *recall*, and *f-value* for the positive prediction results for the classification accuracy. The f-value is calculated as  $\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$ . We apply three-fold cross-validation that uses two parts for training and one part for testing. For random forest, we run the experiment ten times and show the average results.

### 4.3 Results and Discussion

We first test the effectiveness of each individual set of the proposed features, namely, interest keywords, mom-word in tweets, and picture analysis. This result is shown in Table 3. We see that interest keyword provides the highest accuracy with random forest classifier, reaching f-value of 0.579. We also see that the Picture feature set achieves the highest precision with SVM, but has a very low recall. This is reasonable, because if the profile picture contains a baby child, the user would very likely be a baby mother. However, a large portion of baby mothers do not show their babies on profile pictures. Similarly, mom-word features reach a high precision but a low recall with random forest, because if a user uses special words she is likely to be a mother, but not all mothers make mother-related posts.

**Table 3.** Classification accuracy of individual sets of proposed features

	Interest		Mom-word		Picture	
	SVM	RF	SVM	RF	SVM	RF
precision	0.581	0.753	0.443	0.732	<b>0.773</b>	0.728
recall	<b>0.553</b>	0.474	0.471	0.354	0.121	0.095
f-value	0.556	<b>0.579</b>	0.456	0.477	0.209	0.167

Next we compare the proposed feature sets with baseline feature sets, namely BOW, socio-linguistic (SocLing) and postpartum behavior feature (Postpartum). The results are shown in Table 4. In this experiment, the proposed feature set is the combination of all three feature sets discussed above. As we can see from the results, with SVM, the proposed features reach f-value of 0.562, more than 7% higher than those reached by BOW, SocLing, and Postpartum. With random forest, the proposed features reach the highest precision of 0.785, and the highest f-value of 0.623, more than 15% higher compared to those reached by three baselines.

<sup>9</sup> <https://cran.r-project.org/web/packages/e1071/e1071.pdf>.

<sup>10</sup> <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>.

**Table 4.** Classification accuracy comparison of existing and proposed features

	BOW [7]		SocLing [8]		Postpartum [2]		Proposed	
	SVM	RF	SVM	RF	SVM*	RF	SVM	RF
precision	0.471	0.760	0.508	0.761	0.589	0.601	0.587	<b>0.785</b>
recall	0.516	0.343	0.418	0.341	0.228	0.302	<b>0.540</b>	0.519
f-value	0.490	0.472	0.458	0.469	0.326	0.402	0.562	<b>0.623</b>

\*using radial kernel

**Table 5.** Effects of combining existing and proposed features

	BOW + Proposed		SocLing + Proposed		Postpartum + Proposed		All	
	SVM	RF	SVM*	RF	SVM	RF	SVM	RF
precision	0.584	0.790	0.550	0.797	0.786	0.792	0.562	<b>0.808</b>
recall	<b>0.521</b>	0.497	0.520	0.505	0.250	0.501	0.507	0.461
f-value	0.551	0.609	0.534	<b>0.617</b>	0.379	0.613	0.532	0.587

\*using radial kernel

We are also interested in whether combining the proposed features with the established feature sets improves the accuracy. The results are shown in Table 5. Comparing this table to Table 4, we see that by adding the proposed features, all baseline features achieve higher accuracy. It improves f-value of BOW from 0.472 to 0.609, for SocLing from 0.469 to 0.617, and for Postpartum from 0.402 to 0.613. Combining all features together, we achieve a precision of 0.808, the highest among all cases we tested. However, the achieved f-value of 0.587 is lower than that achieved by using the proposed features alone.

## 5 Conclusion

In this paper, we study the problem of finding mothers with baby children on Twitter. Following a supervised machine learning approach, we propose novel features based on followed accounts, vocabulary used, and profile pictures. Experimental results with real Twitter user data show that the proposed features are highly effective for baby mother discovery, and achieve considerably higher classification accuracy compared to three baselines of established feature sets. In future we plan to extend our approach on other user groups such as people with disabilities or rare diseases.

**Acknowledgement.** This research has been supported by JSPS KAKENHI Grant Numbers #17H01828, #18K19841 and by MIC/SCOPE #171507010.



## References

1. Barberá, P.: Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Anal.* **23**(1), 76–91 (2015)
2. De Choudhury, M., Counts, S., Horvitz, E.: Predicting postpartum changes in emotion and behavior via social media. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3267–3276. ACM (2013)
3. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. In: *Proceedings of the Seventh International Conference on Web and Social Media*, vol. 13, pp. 1–10 (2013)
4. MacLean, D., Gupta, S., Lembke, A., Manning, C., Heer, J.: Forum77: an analysis of an online health forum dedicated to addiction recovery. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1511–1526. ACM (2015)
5. Morris, M.R.: Social networking site use by mothers of young children. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1272–1282. ACM (2014)
6. Pennacchiotti, M., Popescu, A.M.: A machine learning approach to Twitter user classification. In: *Proceedings of the Fifth International Conference on Weblogs and Social Media*, pp. 281–288 (2011)
7. Rangel, F., Rosso, P., Potthast, M., Stein, B.: Overview of the 5th author profiling task at pan 2017: gender and language variety identification in Twitter. In: *Working Notes Papers of the CLEF* (2017)
8. Rao, D., Yarowsky, D., Shreevats, A., Gupta, M.: Classifying latent user attributes in Twitter. In: *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents*, pp. 37–44 (2010)
9. Sharma, N.K., Ghosh, S., Benevenuto, F., Ganguly, N., Gummadi, K.: Inferring who-is-who in the Twitter social network. *ACM SIGCOMM Comput. Commun. Rev.* **42**(4), 533–538 (2012)
10. Zhang, Y., Szabo, C., Sheng, Q.Z.: Improved object and event monitoring on Twitter through lexical analysis and user profiling. In: *Proceedings of the 17th International Conference on Web Information System Engineering*, pp. 19–34 (2016)