Zhewei Liu, Anshu Zhang, Yepeng Yao, Wenzhong Shi, Xiao Huang & Xiaoqi Shen (2021) Analysis of the performance and robustness of methods to detect base locations of individuals with geo-tagged social media data, International Journal of Geographical Information Science, 35:3, 609-627, DOI: 10.1080/13658816.2020.1847288

Analysis of the performance and robustness of methods to detect base locations of individuals with geo-tagged social media data

Abstract:

Various methods have been proposed to detect the base locations of individuals, with their geo-tagged social media data. However, a common challenge relating to base-location detection methods (BDMs) is that, the rare availability of ground-truth data impedes the method assessment of accuracy and robustness, thus undermining research validity and reliability. To address this challenge, we collect users' information from unstructured online content, and evaluate both the performance and robustness of BDMs. The evaluation consists of two tasks: the detection of base locations and also the differentiation between local residents and tourists. The results show BDMs are able to achieve high accuracies in base-location detection but tend to overestimate the number of tourists. Evaluation conducted in this study, also shows that BDMs' accuracy is subject to the intensity of user's activities and number of countries visited by the user but are insensitive to user's gender. Temporally, BDMs perform better during weekends and summertime than during other periods, but the best performances appear with datasets that cover the whole time periods (whole day, week and year). To the best of knowledge, this study is the first work to evaluate the performance and robustness of BDMs at individual level.

Key words: base-location detection; geo-tagged social media data; smart tourism

1. Introduction

The rise of Spatial Big Data has provided researchers with new data sources to study various research problems. These new kinds of datasets have been introduced into different fields, to address such issues as measuring economic activity (Dong et al. 2017, Mancini et al. 2018, Sinclair et al. 2018, Sobolevsky et al. 2017), regionalization (Gao et al. 2013, Li et al. 2019, Jia et al. 2019), urban understanding (Zhou et al. 2019, Yao et al. 2019, Zhu et al. 2020) and human mobility (Soundararaj et al. 2020, Yang et al. 2019, Chen et al. 2019).

In the field of smart tourism, a popular application of Spatial Big Data is the detection of people's base locations (i.e., where they currently live) with the geo-tagged social media data (Vu et al. 2015). Previous works have devised several ways for base-location detection, such as detecting the place with most check-ins (Cho et al. 2011), the place with most night-time check-ins(Liu et al. 2018), the most frequent place in the transition trajectory (Huang and Wong 2016, Zhang et al. 2020).

However, one common challenge regarding base-location detection is that, due to the rare availability of users' true base locations at individual level (Vanhoof et al. 2018b), the accuracies of these base-location detection methods (BDMs) have not yet been evaluated, which brings uncertainties into the research methodology and undermines the soundness of relevant research findings. Some previous work compared the results of BDMs with the population density from census and validated the effectiveness of BDMs at aggregate level (Vanhoof et al. 2018b). Nevertheless, such aggregate-level evaluations are basically a rough approximation when the accuracy of BDMs at individual level remain unknown. Additionally, the lack of performance evaluation also limits our understanding of the robustness of BDMs. The sensitivities of BDMs to different model specifications have not yet been revealed. How accurately BDMs perform at individual level and robustness level of the performance of BDMs in different experimental settings, remain unaddressed questions.

Consequently, in this work, to answer the aforementioned questions, we conduct a series of analysis to validate the accuracy of BDMs at individual level and also to evaluate the robustness of BDMs. As mentioned above, one major obstacle hindering the performance analysis of BDMs is that the true base locations of users are rarely available. To tackle this obstacle, we utilize the geo-tagged data from a social media platform where the users are encouraged by the community culture, to leave real information (such as, current living place, gender, education, hobbies). Such personal information left by users are mostly unstructured and jargonistic, hence the information is manually collected and further visually interpreted. Specifically, we target 16 tourist attractions in Hong Kong and retrieve the geotagged posts (both within and outside Hong Kong) of the users who have check-ins at those locations, from Instagram platform. The users' base locations are then detected respectively with several popular BDMs and the results are evaluated against the ground truth collected from users' Instagram homepage profile. The sensitivity of BDMs to several factors are also tested.

Our results indicated that BDMs have good performance (high *Accuracies*, 80%) regarding base-location detection. However, in terms of differentiating local/tourists, BDMs maintain high *Accuracies* (85%) even though there was a high chance (about 40%) of missing local people, and therefore, possibly overestimating the number of tourists and hence introducing inaccuracy into results. In a balanced dataset, the *Accuracies* of BDMs in terms of differentiating local/tourists, would be about 77%. The results also show that the performance of BDMs is affected by the intensity of user activity and count of countries that user has visited, but insensitive to user gender. For temporal sensitivity, BDMs perform better during weekends and summertime than during other periods, however the best performances are achieved with datasets covering the whole time periods (whole day, week and year).

The major contributions of this work are summarized as:

- The accuracy of BDMs at individual level are evaluated and the bias of BDMs is revealed. To the best of our knowledge, this is the first time that BDMs are evaluated at individual level. The findings suggest that BDMs can achieve high accuracy in base-location detection but possibly overestimate tourist numbers. With this work, valuable information can be gained to guide future use of BDMs, by knowing the accuracy and bias of BDMs.

- The robustness and sensitivity of BDMs are analysed. We investigate the performance of BDMs over various factors: (1) demographic features, (2) mobility patterns of users (3) selection of time periods. The relationships between BDMs and various factors are revealed and knowledge is gained about the influences of model specifications on BDMs performance. Suggestions are also made on how to achieve better performance with proper experiment settings.

The rest of the paper is as below. In Section 2, previous related works are reviewed. In Section 3, the key research problems are defined and the research methodology is introduced. In Section 4, the experiment results are given. In Section 5, the implications and findings are discussed. In Section 6, the whole paper is concluded.

2. Related works

To detect the individual base locations with their online footprints, the frequency of users' posts is commonly used to infer their base locations. For example, user's base location can be detected as the place where users post most check-ins (Hawelka et al. 2014, Yuan and Medel 2016, Li and Yang 2017, Sen and Dietz 2019). A revision of this method is to exclusively use posts during night time, such as from 8 pm to 8 am (Liu et al. 2018), as people tend to stay at home during night. The temporal and post count constraints can also be combined, hence detecting base locations as the places with both most posts and longest timespan between the first and last post taken there (Paldino et al. 2016). Another consideration is that the base location is always the place where he/she returns after travelling. Consequently, the user's base location can be detected by removing the repeated place items in the trajectory and identifying the most frequent place (Huang and Wong 2016, Zhang et al. 2020). Cesario et al. (2016) formulates the base-location detection as a sequential pattern mining problem, aiming to find the sequences larger than certain support level. Bojic et al. (2015) divides the base-location detection methods into five categories: (1) maximum number of posts, (2) maximum number of active days, (3) maximum timespan between the first and last photograph, (4) maximum number of posts from 7 pm to 7 am and (5) maximum number of active days from 7 pm to 7 am.

Another related topic is the differentiation of local residents and tourists, given their mobility patterns. This can be achieved by detecting the user's base location. Once the user's base location is detected, he/she is naturally identified as a non-local tourist in other places. Zheng et al. (2012) constructs mobility entropy for each user and classifies the user as tourist if the entropy exceeds the empirical threshold. Chua et al. (2016) jointly use the users' time-zones and proportions of days spent in specific locations as a means of distinguishing locals and tourists. A popular method is that, if a user has posts in some place and the time period between the first and last post is less than one month, then the user is identified as a tourist (García-Palomares et al. 2015, Habib and Krol 2017, Lee and Tsou 2018, Koutras et al. 2019). In Su et al. (2016), three criteria are used together: (1) the active days are less than one month; (2) timespan is less than three months; (3) base location in the user profile is other country.

Selection of datasets may affect the performance of BDMs, since the user demographics across platforms may be different. Aaron and Monica (2018) find, that Facebook, Instagram, and Twitter all have more female users than male users, while the proportion of female users on Facebook is larger than on Instagram and Twitter. The results also show different races have different preferences, e.g., Hispanic people show preferences for Facebook over Instagram and Twitter. Other demographics features such as age, income, or education level also show different patterns across platforms. There are studies comparing performances of different social media platforms related to specific tasks. Silva et al. (2013) finds that, compared with Instagram, Foursquare is more able to indicate users' typical routes. Tenkanen et al. (2017) finds that Instagram clearly performs better than Twitter and Flickr, in indicating park popularity. Salas-Olmedo et al. (2018) uses three data sources to reflect different activities: Twitter for accommodation, Panoramio for sightseeing, Foursquare for consumption.

After base-location detection and tourist differentiation, how to do result validation has been a long inflicting question. A key problem is that the actual base locations of users are rarely available, so that the results of BDMs can hardly be verified. Bojic et al. (2015) compares the radius of gyrations generated by different methods and concludes that different datasets are unequally susceptible to different detection methods. Some works use ground-truth census and do aggregate-level validation. In Vanhoof et al. (2018a), the official census data in Voronoi polygons of the cell tower network is used to evaluate the home detection results. A similar work is done by Zhu et al. (2018), where the ground truth of migration flow is retrieved from a Chinese web service company (Baidu) and used to evaluate the predicted migration flow.

However, these aggregate-level evaluations can only be a rough approximation and in such cases, how accurately the various BDMs perform at individual level remains unknown. Consequently, in this work, we firstly collect the geotagged posts and detect the base locations of individuals with several popular BDMs. The user's true information (i.e., currently living place, gender, etc.) is then, manually collected from the users' homepage profiles. Finally, the performance and robustness of BDMs are evaluated against the ground truth. Compared with previous related works, this paper is the first work to evaluate the accuracies and sensitivities of BDMs, at individual level.

3. Problem definition and methodology

3.1 Key problem definition

For convenience, the key definitions of this study are given firstly (in Table 1) and used consistently throughout this paper.

Table 1. Key Definitions

Variable	Description
u_i	A unique user in the social media dataset
p_i	A social media post with identifier, user, timestamp, location $p_i = (id, u,\ t,\ loc)$
R	A region is an administrative region (e.g., city, province or country)
UP_i	A user-posting-history, i.e., all the posts in the dataset of a user u_i, $UP_i = (up_1^i, up_2^i, \dots, up_m^i)$

Definition 1: User. A u_i is a unique user in the social media dataset.

Definition 2: Post. A social media post p_i is a record that a user *u* visits location *loc* at timestamp *t*, with a unique identifier, $p_i = (id, u,\ t,\ loc)$

Definition 3: Region. A region *R* is an administrative region (e.g., city, province or country).

Definition 4: User Posting History. A user-posting-history UP_i is defined as all the posts of u_i, $UP_i = (up_1^i, up_2^i, \ldots, up_m^i)$, where up_k^i is the k^{th} post of user u_i.

The base-location detection and recognizing tourist problems in this work can therefore be defined as below:

Base-location Detection Problem: Given a user-posting-history UP_i, we aim to identify a region *R* that is the base location of u_i.

Recognizing Tourist Problem: Given a region *R* and user-posting-history UP_i, we aim to identify whether u_i's base location is *R* or not.

3.2 Base-location detection methods (BDMs)

In this work, several popular base-location detection methods (BDMs) are implemented and their performances evaluated. These methods are commonly used in previous works (Cho et al. 2011, Bojic et al. 2015, Huang and Wong 2016, Yuan and Med el 2016, Liu et al. 2018, Vanhoof et al. 2018a, Sen and Dietz 2019, Zhang et al. 2020). The details of the implemented methods are given below:

1. The base location is detected as the place where a social media user makes maximal posts. (MP)

2. The base location is detected as the place where a social media user spends the maximum number of active days, where an active day is a day when a user makes one or more posts. (MAD)

3. The base location is detected as the place where a social media user makes maximal posts from 7 PM to 7 AM. (MP-19-7)

4. The base location is detected as the place where a social media user spends the maximal number of active days from 7 PM to 7 AM. (MAD-19-7)

5. The base location is detected as the place with the most frequency of visiting, after removing the repeated place entity in user's posting history. As described in Section 3.1, the posting history of a user u_i is defined as $UP_i = (\ up_1^i, up_2^i, \ldots, up_m^i)$, and $up_k^i.loc$ is the place where user u_i makes his k^{th} post. After removing the repeated sequential place entities in PH, the most frequently visited place is identified as user's base location. (MFV)

In this work, we study base-location detection methods (BDMs) in the country level. The base location of a user is defined as the country where the user currently lives.

3.3 Evaluation of performance and sensitivity

In this work, two tasks of BDMs are investigated: (1) the detection of a user's base location, (2) the identification of a user as local/tourist. For base-location detection, we use the measurement of *Accuracy*#, which is the fraction of correct predictions among all predictions. For identifying local/tourist, another four measurements are introduced, i.e., *Precision*, *Recall* (*true positive rate*), *Specificity* (*true negative rate*) and *Balanced Accuracy*. *Precision* is the fraction of correction positive predictions among all positive predictions. *Recall* is the fraction of correct positive predictions among all positive instances. *Specificity* is the fraction of correct negative predictions among all negative instances. *Balanced Accuracy* is an overall performance measurement for imbalanced datasets, calculated by averaging *Recall* (*true positive rate*) and *Specificity* (*true negative rate*). In our experiments, locals are identified as positive and non-locals (tourists) as negative.

In addition, we also do sensitivity analysis, to investigate how the following factors: (1) demographic features of users, (2) mobility patterns of users and (3) selection of time periods, affect the performance of BDMs.

(Note#: In this paper, "Accuracy" refers to this specific measurement, while "accuracy" is a broader concept, referring to how well BDMs perform generally.)

4. Experiment and results

In this section, the datasets are first described, and followed by the experiment results.

4.1 Experiment settings

Geo-tagged social media check-ins from Instagram are used in the experiments. The data is collected using Instagram APIs. The data collection is a two-step process. First, we target 16 popular tourist attractions in Old Town Central (OTC), Hong Kong, and collect the Instagram check-ins recorded in these 16 attractions (shown in Table 2). OTC is one of the oldest and also most dynamic districts in the city. The diverse nature of the 16 target sites attracts both local residents and tourists from all over the world. Secondly, the Instagram users of these collected check-ins are identified and all their historical geo-tagged check-ins (regardless of whether in Hong Kong or not) are collected.

Table 2. Targeted attractions in OTC, Hong Kong

Category	Attraction
Modern attractions	PMQ, Tai Kwun, Pottinger Street, Hollywood Road, Gough Street, Museum of Medical Sciences, Possession Street, Yan Gallery
Historical attractions	Man Mo Temple, YMCA Bridges Street Centre
Art attractions	Fringe Club, La Galerie - Paris 1839, Karin Weber Gallery
Urban attractions	Parkview ART, Pak Tsz Lane Park, Tai Ping Shan Street

The Instagram culture encourages users to write their real information in their homepages (Figure 1). These kinds of user information are publicly accessible yet very unstructured, hence making it difficult to automatically extract user information with programming or natural language processing tools. Consequently, we go to each user's homepage and manually collect their information, one by one. As shown in Figure 1, we visually interpret each user's homepage and manually extract relevant information (e.g., currently living place, gender, education, hobby etc.). Each user's identity is also checked. Commercial accounts such as those that represent shops or companies and the users without information of currently living places are removed.

Key fields of the geo-tagged check-ins related to the above experiments are listed in Table 3.

Figure 1. One Instagram user homepage profile, with information such as currently living place, gender, education, hobby

Table 3. Fields of geo-tagged check-ins from Instagram

Fields	**Description**
cid	A string uniquely indicating the check-in
user_name	A string uniquely indicating the user
gender	The gender of the user (Female, Male or Unknown)
time	The timestamp when the check-in is posted
location	The location where the check-in is posted

After data cleaning, 821,331 check-ins generated 2159 users are recorded in total. The ground-truth base locations of these users are shown in Figure 2.

Figure 2. Distribution of actual base locations of users recorded in this work

4.2 Results of performance evaluation

Five BDMs are implemented with the above datasets and each user's base location is detected. We measure the *Accuracy* of methods against the ground truth in terms of detecting user's base location. Results are given in Table 4. All five BDMs provided *Accuracies* of around 80%, indicating that most users' base locations could be correctly detected. Among all BDMs, MAD achieves the best performance (84.5% *Accuracy*), while MFV achieves the worst performance (75.4% *Accuracy*). It also shows MP and MAD could both achieve a better performance, than their counterparts that only take night-time check-ins (MP-19-7 and MAD-19-7). This finding challenges the previous assumption that night-time check-ins are more indicative of user's base location.

Table 4. *Accuracies* of BDMs in terms of user base-location detection

BDMs	**MP**	**MAD**	**MP-19-7**	**MAD-19-7**	**MFV**
Accuracy	82.5%	84.5%	76.3%	78.4%	75.4%

The performances of BDMs in terms of identifying local/tourists are further evaluated. As the experiment datasets are collected from users who have visited specific tourist attractions in Hong Kong, Hong Kong residents are regarded as a) local (positive) and b) tourists to Hong Kong as non-local (negative). Measurements of *Precision*, *Recall*, *Specificity* and *Accuracy* (in Section 3.3) are calculated (shown in Figure 3). Similar to our previous finding, the MP and MAD outperforms their counterparts (MP-19-7 and MAD-19-7), when measured by *Precision* and *Accuracy*, yet all methods performed well , each with both measurements over 85%. For *Specificity*, the scores are all over 95%, indicating that majority of the non-local tourists can be correctly predicted as tourists by BDMs.

Figure 3. *Precisions, Recalls, Specificities, Accuracies* of BDMs in terms of identifying local/tourist

However, the measurement *Recalls* for all BDMs are relatively low, i.e., around 60%, meaning that nearly half of local people are not correctly identified, hence indicating that BDMs have a high chance of missing positive instances (local people). The results possibly indicate that BDMs could be over strict regarding their identification of local people: even though most users identified as local are correct (averagely 90% of those who are identified as local are actual local), there are some (40%) actual local people misidentified as tourists.

4.3 Influence of the Imbalanced Dataset

BDMs maintain high *Accuracies* (around 85%), even they have high chances (about 40%) of missing positive instances (local people). This is because local people account for only 27% of the total population in the datasets (Figure 2). For such imbalanced datasets, *Accuracy* can be a misleading measurement.

Consequently, to correct the influence of imbalanced datasets, we introduce another measurement *Balanced Accuracy*, which is widely used for measuring the overall performance of a model on imbalanced datasets (García et al. 2009, Brodersen et al. 2010). As described in Section 3.3, *Balanced Accuracy* is calculated by averaging the sum of *Recall* (true positive rate) and *Specificity* (true negative rate):

$$Balanced\ \ Accuracy = \frac{Recall\ \ +\ \ Specificity}{2} \tag{1}$$

For an imbalanced dataset (e.g., 90 negative and 10 positive instances), predicting all as negative will achieve 0.5 *Balanced Accuracy* score, same as the expected value of a random guess in a balanced dataset.

Table 5. *Balanced Accuracies* of BDMs in terms of identifying local/tourist

BDMs	MP	MAD	MP-19-7	MAD-19-7	MFV
Balanced Accuracy	76.4%	79.1%	76.0%	77.5%	80.4%

The calculated *Balanced Accuracies* are shown in Table 5. The Table shows that *Balanced Accuracies* achieved by BDMs are about **77%**. Such results indicate that if the dataset is balanced, i.e., the number of positive instances (local people) is the same as negative instances (non-local tourists), then, among every ten predictions, averagely there should be about eight correct predictions. On the other hand, for an imbalanced dataset, if we assume the importance of true positive rate and true negative rate are the same, the overall performance of BDMs should then also fall into this range.

We further quantify the relationships between the overall *Accuracy* and the degree of dataset's imbalance. The formula is given below:

$$Accuracy = \frac{correct\ \ positive\ \ predictions + correct\ \ negative\ \ predictions}{all\ \ instances}$$

$$= \frac{correct\ \ positive\ \ predictions}{positive\ \ instances} * \frac{positive\ \ instances}{all\ \ instances} + \frac{correct\ \ negative\ \ predictions}{negative\ \ instances} * \frac{negative\ \ instances}{all\ \ instances}$$

$$= Recall * r + Specificity * (1 - r) \tag{2}$$

where *r* is the ratio of positive instances among all instances, i.e., in our case, the percentage of local residents among the total population in the dataset. From this formula (2), it can be seen that given certain *Recall* and *Specificity*, the overall *Accuracy* of the methods is a linear function of the degree of dataset's imbalance. The formula (2) also gives the upper and lower bounds of overall *Accuracy*:

$$Accuracy \in \begin{cases} [Recall, Specificity], \text{if } \ Recall \ < \ Specificity \\ [Specificity, Recall], \qquad\qquad \text{otherwise} \end{cases} \tag{3}$$

In our case, it can be estimated that if the percentage of local people in the datasets increases, the overall *Accuracy* will drop accordingly. By giving formula (2), the relationship between the overall *Accuracy* and the degree of the dataset's imbalance is quantified.

4.4 Results of sensitivity evaluation

Performances of BDMs may be affected by (1) users' demographic features (2) mobility patterns and (3) selection of time periods. Consequently, in this section, several kinds of sensitivity analysis are conducted, with the intention of answering the following questions:

- **Gender Difference**: Female users are, on average, more active on social media, than male users. Will this affect the performances of BDMs across different genders (male and female)? Is gender a factor affecting BDMs?
- **User Behavior Patterns**: Whether BDMs perform better with the users who are more active in the location-based social networks (LBSN)? Will the base locations of people with wanderlust be more unpredictive, since they are regularly traveling?
- **Temporal Difference**: Is there any time period (e.g., month or day of the week) during which BDMs perform better than that during other periods?

By measuring the performance in different scenarios, BDMs' sensitivity to above factors will be evaluated.

4.4.1 Sensitivity to gender difference

Out of 2159 users depicted in the datasets, there are 1169 female users and 718 male users, while other users' genders remain unknown. We evaluate BDMs' performance across female and male users (shown in Figure 4). Performances for different genders are similar to the average performance shown in Table 4 and Figure 3. For both tasks (base-location detection and identifying local/tourist), *Precisions* and *Accuracies* maintain high value (around 85%), with *Recalls* around 60%. The performance differences across female and male users are all less than 5% (mostly less than 3%), indicating that BDMs' performances show no significant difference across different genders.

(a) Performance difference in terms of base-location detection (b) Performance difference in terms of identifying local/tourist

Figure 4. BDMs' performance for different genders. (a) Performance of BDMs (*Accuracy*) across female and male users in terms of base-location detection. (b) Performance of BDMs (*Precision*, *Recall*, *Accuracy*) across female and male users in terms of identifying local/tourist.

4.4.2 Sensitivity to user behaviour patterns

In this section, we investigate the following issues: (1) Whether the base locations of active users in LBSN are easier to detect? (2) Whether the base locations of the people with wanderlust will be more unpredictable? Two measurements are introduced: (1) post count of a user, to quantify user's intensity of activity; (2) count of the countries visited by a user to quantify the user's wanderlust. The base-location detection results with BDMs are transformed into binary value: if the detected base location is correct, then it is assigned as 1, otherwise as 0.

Table 6. The strength of correlation between variables (Point-Biserial correlation coefficient; * Correlation is significant at 0.05 level, $p < 0.05$; ** Correlation is significant at 0.01 level, $p < 0.01$)

BDMs	MP	MAD	MP-19-7	MAD-19-7	MFV
Post count of a user	0.039	0.041	0.061**	0.062**	-0.035
Count of the countries a user has visited	-0.058**	-0.051*	-0.076**	-0.072**	-0.095**

Point-Biserial Correlation Coefficient is calculated to measure the strength of correlation (shown in Table 6). The results reveal that for MP-19-7 and MAD-19-7, user's intensity of activity in LBSN and chances that user's base location is correctly detected are positively correlated. The correlation coefficients are 0.061 and 0.062 (weakly positively correlated), with the significance at 0.01 level. While the results for MP, MAD, MFV, show no correlation with user's intensity of activity at 0.05 level. This indicates that, for MP-19-7 and MAD-19-7, the more a user post his/her footprints in LBSN, the more likely his/her actual base location can be correctly detected. On the other hand, for all the five BDMs, the count of the countries that a user has visited are negatively correlated. The values of correlation coefficients for BDMs range from [-0.095, -0.051] (weakly negatively correlated), with MAD significant at 0.05 level and other methods significant at 0.01 level.

The above findings support the following conclusions: (1) Among the five BDMs, MP-19-7 and MAD-19-7 perform better with the users who are more active, i.e., the more a user post on LBSN, the more likely his/her base location will be corrected detected. (2) For all the BDMs, the more countries a user visits, the less likely that his/her actual base location will be correctly detected, i.e., base locations of the people with wanderlust are more unpredictable, since they are regularly traveling, rather than settling in one place for long.

4.4.3 Sensitivity to temporal difference

In this section, we investigate the question whether BDMs perform better over specific time periods than other periods. We select posts within specific time periods (different months and days of the week) and calculate their corresponding *Accuracies* of BDMs in terms of base-location detection. For monthly performance (shown in Figure 5), results show that all the BDMs display similar trends and perform better during summertime (May to September). Their performances improve gradually from January to May, and remain at relatively high levels from May to September, then decrease in October and remain stable in November and December. MP, MAD, MAD-19-7, MFV all achieve the best performance in August, while MP-19-7 achieves its best performance in June.

Figure 5. *Accuracies* of BDMs in terms of base-location detection over different months

Weekly performance (Figure 6), shows that all the BDMs exhibit similar trends. Their performances decrease gradually from Monday and bottom on Wednesday or Thursday, then bounce back and achieve the best performances on weekends. An explanation is that, on weekdays, people are probably working and on business to other places, while on weekends, they return home for leisure. Such working and living patterns cause the fluctuation of BDMs' performances from Monday to Sunday.

Figure 6. *Accuracies* of BDMs in terms of base-location detection over different days of the week

Another finding concerns the data bias caused by the selection of specific time periods. This finding is achieved by examining the gaps between the original performance of BDMs (Table 4) and their performances over specific periods (Figure 5 and Figure 6). On average, by using datasets from only one specific month for experiments will cause BDMs' performances to drop by 20%-30%, compared with figures when using the datasets covering the whole year. Similarly, using datasets from only one day of the week will cause the performances drop by 8%-15%. From Table 4 and Figure 3, it is also seen that using datasets from night time only, will also cause performance to drop, compared with results, using datasets for the whole day. The above results suggest that, even although BDMs, may perform better during specific time periods than during other periods, datasets only covering specific time periods will inevitably cause data bias and thus underperformance. Only by using datasets covering the whole time periods (whole day, week and year), can BDMs achieve their best performances.

5. Discussion and implications

In this section, the implications of previous findings are summarized. The bias and inaccuracy of BDMs are discussed. The generalizability of the results is given. Finally, the limitations and potential future work are listed.

5.1 Performance and sensitivity of BDMs

Performances of BDMs are evaluated in terms of two tasks: detecting user's base location and identifying users as local/tourist. For both tasks, the average *Accuracies* of the five investigated methods are about 80%. Besides, for identifying local/tourist, the average *Precision* of BDMs is over 85%, meaning that 9 out of 10 users identified as local are actually local. These results show that BDMs investigated in this work have good performance in terms of the abovementioned two tasks.

The sensitivity analysis finds that (1) performances of BDMs across male and female users are basically the same, indicating that BDMs are insensitive to user's gender; (2) for MP-19-7 and MAD-19-7, the more a user posts on LBSN, the easier his/her base location be correctly detected. (3) for all the five BDMs, the more countries a user has visited, the harder his/her base location be corrected detected.

5.2 Bias and inaccuracy of BDMs

Identifying users as local/tourist with BDMs suffers from certain bias. The *Recalls* of BDMs (Figure 3) are around 60%, meaning that 4 out of 10 genuine local residents will be misidentified as tourists. This could indicate that BDMs are over strict in identifying local people: even most users identified as local are correct, some (40%) genuine local people are missed. This finding suggests that previous works using BDMs may underestimate the number of local residents and overestimate the number of tourists, hence bringing potential bias into conclusions.

Another finding is that BDMs can maintain high *Accuracies* (about 85%) even they have high chances of missing local residents. This is because the misidentified users are mainly local residents, misidentified as tourists and local residents actually make up a small portion (27%) in the total datasets. It can be expected that the overall *Accuracies* of BDMs will drop when the portion of local people increase. In this study, the base locations of non-local tourists are identified at country level, while the base location of local residents is identified at a more granular level (i.e., Hong Kong). This may explain the difference of *Accuracies* across local/tourist (60% *Accuracy*/*Recall* for local resident versus over 90% *Accuracy* for tourists).

The experiments also show that inaccuracy may be introduced by selecting datasets from specific time periods for experiments. Although BDMs may perform better during specific time periods (e.g., summertime, weekend) than during other periods, the best performances of BDMs are achieved when datasets cover the whole time period (whole day long, week and year).

5.3 Generalizability of the Results

Different social media datasets may cover different user groups, which may thus affect the generalizability of our results. Nevertheless, we argue certain conclusions revealed in this study can be further extended to other datasets.

Firstly, to calculate the accuracy of BDMs in terms of differentiating tourists from locals, a generalized formula is given in formula (2). This formula shows that, for any dataset, the aforementioned accuracy is determined by three variables: *Recall*, *Specificity* and r (ration of positive instances among all instances). Across different datasets, these three variables may vary, while this mathematical relationship from these variables to the accuracy remains constant. This given formula is universally extensible to other datasets.

Secondly, several conclusions of the robustness evaluation are generalizable. These include: (1) BDMs perform better for active users, as active users tend to post more footprints; (2) the more countries a user visits, the harder it is to correctly detect the user's base location, as users with wanderlust are regularly travelling rather than staying in one place for long; (3)

Weekend check-ins are more indicative of users' base locations than weekday check-ins, as people tend to stay at home more on weekends than weekdays. These conclusions result from people's constant movement patterns and are consistent with common sense. Regardless of the platforms, these constant patterns hold true. We thus argue that above conclusions are extensible to other datasets.

On the other hand, the performance of BDMs may vary across different datasets, since the covered user groups may be different. A thorough study of user demographics and posting patterns across different platforms, may be helpful to identify how well BDMs perform with such various datasets.

5.4 Limitation and future work

There are several worthy questions that can serve as future directions for work related to this paper's topic. Firstly, detecting user's base location in this work is concerned mainly with detecting the country where the user is currently living. Whether base locations of users can be detected at a more detailed level (e.g., province and city level) and how BDMs perform in these levels can present a potential direction. Secondly, the current BDMs are statistical models based on human intuition. These BDMs are easy to understand but have presented an unsatisfactory performance in certain situations. Some other algorithms (e.g., machine learning) may be introduced to improve the performance of BDMs. Thirdly, the current BDMs mainly use the sequence of geo-tagged check-ins to detect user's base location. However, other forms of information such as textual and visual information (the words and figures that the user posts) may also be used to detect user's base location. Fourthly, different datasets may cover different user groups, which may affect the accuracy of BDMs. How BDMs perform with various datasets and to what degree the evaluation results can be generalizable, remain challenging yet significant topics. Fifthly, this study collected users' base locations from their social media profiles, in which users may post fake or out-of-date information. The veracity of this data source needs to be further investigated.

6. Conclusions

Various kinds of methods have been developed to detect users' base locations. However, a challenge is that, due to the rare availability of users' real base locations, the accuracy of these methods has yet to be fully evaluated.

In this work, we manually collect users' real information from unstructured user-generated content and evaluate the performance and robustness of various BDMs with geo-tagged social media data. Our results show that BDMs achieve high *Accuracies* in the task of base-location detection. In terms of differentiating local/tourist, BDMs achieve high *Precisions* and *Accuracies*, yet relatively low *Recalls*, indicating that BDMs may overestimate the number of non-local tourists and thus bring potential bias into research conclusions. Regarding robustness analysis, the results show that the performance of BDMs is affected by the intensity of user activity and number of countries visited by them, while insensitive to user gender. Moreover, regarding temporal sensitivity, BDMs perform better during weekends and summertime than during any other periods, however, the best performances are achieved when datasets cover whole time periods (such as whole day, week and year).

To the best of our knowledge, this study is the first time that performances and robustness of BDMs are evaluated at individual level. Valuable information can be obtained to benefit future use of BDMs, by knowing how accurately BDMs perform, where the bias of BDMs lies and how to achieve better performance with proper experiment settings.

Data and codes availability statement

The data and codes that support the findings of this study are available with a DOI at https://doi.org/10.6084/m9.figshare.12362567

Acknowledgments

We would like to express our gratitude to editor Prof. May Yuan and Dr Michela Bertolotto and the anonymous reviewers, for their insightful comments and feedback, especially during all the chaos caused by Covid-19.

Funding

This work was supported by The Hong Kong Polytechnic University [1-ZVN6]; the Ministry of Science and Technology of the People's Republic of China [2019YFB2103102].

Disclosure statement

No potential conflict of interest was reported by the authors.

References:

Aaron, S. and Monica, A., 2018. *Social Media Use in 2018* [online]. Available from: https://www.pewresearch.org/internet/2018/03/01/social-media-use-2018-appendix-a-detailed-table/ [Accessed August 11 2020].

Bojic, I., *et al.*, Choosing the right home location definition method for the given dataset. ed. *International Conference on Social Informatics*, 2015, 194-208.

Brodersen, K. H., *et al.*, The balanced accuracy and its posterior distribution. ed. *2010 20th International Conference on Pattern Recognition*, 2010, 3121-3124.

Cesario, E., *et al.*, Analyzing social media data to discover mobility patterns at EXPO 2015: Methodology and results. ed. *2016 International Conference on High Performance Computing & Simulation (HPCS)*, 2016, 230-237.

Chen, P., *et al.* 2019. STLP-GSM: a method to predict future locations of individuals based on geotagged social media data. *International Journal of Geographical Information Science,* 33(12), 2337-2362.

Cho, E., Myers, S. A. and Leskovec, J., Friendship and mobility: user movement in location-based social networks. ed. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, 1082-1090.

Chua, A., *et al.* 2016. Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management,* 57, 295-310.

Dong, L., *et al.* 2017. Measuring economic activity in China with mobile big data. *EPJ Data Science,* 6(1), 29.

Gao, S., *et al.* 2013. Discovering Spatial Interaction Communities from Mobile Phone D ata. *Transactions in GIS,* 17(3), 463-481.

García-Palomares, J. C., Gutiérrez, J. and Mínguez, C. 2015. Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography,* 63, 408-417.

García, V., Mollineda, R. A. and Sánchez, J. S., Index of balanced accuracy: A performance measure for skewed class distributions. ed. *Iberian conference on pattern recognition and image analysis*, 2009, 441-448.

Habib, M. B. and Krol, N. C., What Does Twitter Tell Us About Tourists' Mobility Behavior? A Case Study on Tourists in The Netherlands and Belgium. ed. *PACIS*, 2017, 34.

Hawelka, B.*, et al.* 2014. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science,* 41(3), 260-271.

Huang, Q. and Wong, D. W. S. 2016. Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? *International Journal of Geographical Information Science,* 30(9), 1873-1898.

Jia, T.*, et al.* 2019. Detecting the regional delineation from a network of social media user interactions with spatial constraint: A case study of Shenzhen, China. *Physica A: Statistical Mechanics and its Applications,* 531, 121719.

Koutras, A., Nikas, I. A. and Panagopoulos, A., 2019. Towards Developing Smart Cities: Evidence from GIS Analysis on Tourists' Behavior Using Social Network Data in the City of Athens. *Smart Tourism as a Driver for Culture and Sustainability.* Springer, 407-418.

Lee, J. Y. and Tsou, M.-H., Mapping spatiotemporal tourist behaviors and hotspots through location-based photo-sharing service (Flickr) data. ed. *LBS 2018: 14th International Conference on Location Based Services*, 2018, 315-334.

Li, D. and Yang, Y., 2017. GIS monitoring of traveler flows based on big data. *Analytics in smart tourism design.* Springer, 111-126.

Li, Y., Fei, T. and Zhang, F. 2019. A regionalization method for clustering and partitioning based on trajectories from NLP perspective. *International Journal of Geographical Information Science,* 33(12), 2385-2405.

Liu, Q., Wang, Z. and Ye, X. 2018. Comparing mobility patterns between residents and visitors using geo‐tagged social media data. *Transactions in GIS,* 22(6), 1372-1389.

Mancini, F., Coghill, G. M. and Lusseau, D. 2018. Using social media to quantify spatial and temporal dynamics of nature-based recreational activities. *PLoS One,* 13(7).

Paldino, S.*, et al.* 2016. Uncovering urban temporal patterns from geo-tagged photography. *PLoS One,* 11(12).

Salas-Olmedo, M. H.*, et al.* 2018. Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management,* 66, 13-25.

Sen, A. and Dietz, L. W. 2019. Identifying travel regions using location-based social network check-in data. *Frontiers in Big Data,* 2, 12.

Silva, T. H.*, et al.*, A comparison of Foursquare and Instagram to the study of city dynamics and urban social behavior. ed. *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, 2013, 1-8.

Sinclair, M., Ghermandi, A. and Sheela, A. M. 2018. A crowdsourced valuation of recreational ecosystem services using social media data: An application to a tropical wetland in India. *Science of the Total Environment,* 642, 356-365.

Sobolevsky, S.*, et al.*, Predicting regional economic indices using big data of individual bank card transactions. ed. *2017 IEEE International Conference on Big Data (Big Data)*, 2017, 1313-1318.

Soundararaj, B., Cheshire, J. and Longley, P. 2020. Estimating real-time high-street footfall from Wi-Fi probe requests. *International Journal of Geographical Information Science,* 34(2), 325-343.

Su, S.*, et al.* 2016. Characterizing geographical preferences of international tourists and the local influential factors in China using geo-tagged photos on social media. *Applied Geography,* 73, 26-37.

Tenkanen, H.*, et al.* 2017. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific reports,* 7(1), 1-11.

Vanhoof, M., Lee, C. and Smoreda, Z. 2018a. Performance and sensitivities of home detection from mobile phone data. *arXiv preprint arXiv:1809.09911*.

Vanhoof, M.*, et al.* 2018b. Assessing the quality of home detection from mobile phone data for official statistics. *Journal of Official Statistics,* 34(4), 935-960.

Vu, H. Q.*, et al.* 2015. Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management,* 46, 222-232.

Yang, X.*, et al.* 2019. Spatial heterogeneity in spatial interaction of human movements—Insights from large-scale mobile positioning data. *Journal of Transport Geography,* 78, 29-40.

Yao, Y.*, et al.* 2019. A human-machine adversarial scoring framework for urban perception assessment using street-view images. *International Journal of Geographical Information Science,* 33(12), 2363-2384.

Yuan, Y. and Medel, M. 2016. Characterizing international travel behavior from geotagged photos: A case study of flickr. *PLoS One,* 11(5).

Zhang, F.*, et al.* 2020. Uncovering inconspicuous places using social media check-ins and street view images. *Computers, Environment and Urban Systems,* 81, 101478.

Zheng, Y.-T., Zha, Z.-J. and Chua, T.-S. 2012. Mining travel patterns from geotagged photos. *ACM Transactions on Intelligent Systems and Technology (TIST),* 3(3), 1-18.

Zhou, B.*, et al.* 2019. Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision,* 127(3), 302-321.

Zhu, D.*, et al.* 2018. Inferring spatial interaction patterns from sequential snapshots of spatial distributions. *International Journal of Geographical Information Science,* 32(4), 783-805.

Zhu, D.*, et al.* 2020. Understanding place characteristics in geographic contexts through graph convolutional neural networks. *Annals of the American Association of Geographers,* 110(2), 408-420.