The Impact of Biases in the Crowdsourced Trajectories on the Output of Data Mining Processes

Anahid Basiri
Centre for Advanced Spatial Analysis,
University College London
Gower Street
London, the UK
A.basiri@ucl.ac.uk

Muki Haklay
Department of Geography,
University College London
Gower Street
London, the UK
M.haklay@ucl.ac.uk

Zoe Gardner
Nottingham Geospatial Institute
University of Nottingham
Triumph Road
Nottingham, UK
Zoe.gardner@nottingham.ac.uk

Abstract

The emergence of the Geoweb has provided an unprecedented capacity for generating and sharing digital content by professional and non-professional participants in the form of crowdsourcing projects, such as OpenStreetMap (OSM) or Wikimapia. Despite the success of such projects, the impacts of the inherent biases within the 'crowd' and/or the 'crowdsourced' data it produces are not well explored. In this paper we examine the impact of biased trajectory data on the output of spatio-temporal data mining process. To do so, an experiment was conducted. The biases are intentionally added to the input data; i.e. the input trajectories were divided into two sets of training and control datasets but not randomly (as opposed to the data mining procedures). They are divided by time of day and week, weather conditions, contributors' gender and spatial and temporal density of trajectory in 1km grids. The accuracy of the predictive models are then measured (both for training and control data) and biases gradually moderated to see how the accuracy of the very same model is changing with respect to the biased input data. We show that the same data mining technique yields different results in terms of the nature of the clusters and identified attributes.

Keywords: Bias, Crowdsourcing, Trajectory, Spatio-temporal Data Mining

1 Introduction

The Geoweb (Haklay et al., 2008) is now taken for granted as part of the digital infrastructure of everyday modern life. Many applications and services directly benefit from user-generated content contributed by a wide range of users through crowdsourcing projects. Over the past decade, it has become possible for a much wider group of contributors to create and share geographical information.

Despite the success of many crowdsourcing projects, the impact of their inherent demographic biases (e.g. gender, educational and socio-economic background), as well as biases in temporal, spatial, and thematic components of the contributions on the data are not well understood. These biases, which can ultimately determine the characteristics of spatial data, are inevitable in many crowdsourcing projects and are integral to the information that is held by them (Bell et al., 2015; Salk et al., 2016; Brown, 2017).

While quality issues of crowd-sourced data have been studied widely, the identification and estimation of biases in crowd-sourced projects have not received the same attention. This is due mainly to the lack of availability of the (geo-) demographic data of the contributors, which is either unrecorded (e.g. OpenStreetMap - OSM) or inaccessible due to data protection (e.g. Google MapMaker). Therefore understanding the impacts of demographic biases on crowdsourced maps is challenged by a lack of data on these aspects. Mullen et al. (2015) evaluated the impacts of (geo-) demographic characteristics on the spatial accuracy in two Volunteer Geographic Information (VGI) projects. However, the study was limited by the inference of the demographic structure of the population by local census data rather than recorded demographic data. A further study by Gardner et al., (2018) built on this approach but instead collected demographic data from individual OSM users to explore the impact of (geo-) demographics on VGI (in this case gender), conducted in the context of OSM. It revealed that men have a higher propensity than women to modify existing data as well as demonstrating more variance in their preferences for feature tagging. These observed differences in gendered crowdsourced mapping could impact on positional and thematic accuracy respectively (Gardner et al., 2018). Further work based on the same data supports this proposition. Using a small sample of users' edits to the Malawian national OSM dataset, Gardner and Mooney (2018) found that men placed a much greater emphasis on the geometric accuracy of their edits than female editors. However, comparison to the authoritative dataset would be required to test this assertion.

Crowdsourced and open data have brought an unprecedented opportunity to researchers to analyse and extract knowledge, particularly using statistical and machine learning (ML) techniques (Basiri et al., 2016a). Projects such as OSM now share over 300 Gigabytes of traces of movements using Global Positioning System (GPS) receivers in mobile devices. The maturity of this area means that rich information are available for further analysis, providing both longevity and spatial coverage that can allow change detection, event identification, and the extraction of meaningful information from the pattern of adding features (Neis et al., 2012) and tags over time and space (Mooney and Corcoran, 2014). However applying statistical and machine learning techniques to a potentially biased dataset may result in recognising patterns for a minority of users or ignoring some existing trends (Mooney, 1996).

Many machine learning techniques are too complex to decompose bias and variance. Also in many cases the quality and biases in the data are insufficiently understood to use conventional de-biasing techniques, such as sample re-weighting (Howard et al., 2017). Also, due to the nature of VGI, "Bootstrap" sampling may not practically provide a good measure for biases (Hinds, 1999). Finally, and more importantly, many crowdsourcing projects tend to protect the privacy of their contributors and so it is important to know how the biased dataset, as is, may fit the purpose of using each machine learning technique. Our current study addresses the impacts of using biased data on the results of the analysis. This may help to have a better understanding of the accuracy, reliability and fitness-for-purpose of the outputs.

This paper is structured as follows; The following second section reviews the biases and discusses some of the common types of biases that exist in volunteered/crowdsourced geographic data. The third section explains the methodology and experiments and discusses the results.

2 Bias in VGI

The purpose of many machine learning techniques is to find trends and patterns within data. This may mean classifying and clustering data as well as labelling them. While this process may appear to 'stereotype' some of the observations and therefore neglect minorities and micro-patterns, the underlying data must be as un/de-biased as possible. This is mainly due to the learning methods that both humans and machines share to some extent, i.e. learning from the past. Mitchell (1980), Schaffer (1994) and Wolpert (1996) showed that bias-free learning is futile. Therefore recognising valuable insights from (potentially complex) datasets could be vulnerable to data anomalies, errors and biases as biased training data may potentially send algorithms astray and make "the winners always win". The inference and knowledge extraction are the ultimate goal of any statistical and machine learning algorithm and according to Sackett (1979) therefore any biases in such judgmental processes need to be well studied. This section focuses on the sources of biases in voluntarily contributed data and the process of knowledge extraction from it.

Any observation, and therefore any data contributed through VGI projects, is prone to random and/or systematic errors. While random error can be reduced by increasing the sample size, the systematic errors are more to do with design, methodology and other procedures that lead in obtaining data with no significant correlation with the sample size. Any VGI project is biased in one or more ways. At the first glance, it seems that all the data contributed through VGI projects are "voluntary response samples", which are always biased as they only include people who have chosen to volunteer (DeMaio, 1980). Whereas a random sample would need to include people whether or not they choose to volunteer (Goyder, 1986). Thus inferences from a voluntary response sample are not as trustworthy as conclusions are based on a random sample of the entire population. While crowdsourcing projects are technically open to the whole population, and of course, anyone should be able to contribute, recent studies (Mullen et al., 2015; Gardner et. al. 2018; Zhu et al., 2017) have shown that even the most popular crowdsourced projects, such as OSM, are biased by the contribution patterns of its contributors, i.e. that a small percentage of the community contribute the greatest proportion of activity (the 'long tail effect' or 90-9-1 rule). This therefore questions the use of the terms "crowd" and "public" used in many crowdsourcing and public participatory projects by virtue of this skewed pattern of participation. This excludes the projects which may require a relatively higher experience level, access to some resources, or may limit participation to a specific geography or particular time interval due to the nature of the project.

In addition to voluntary response bias, the volunteers, as individuals, can have different aspects and levels of quality of judgement and decision making (Hammond, 2000). Their decisions, opinions, and preferences could be significantly represented and/or influence their contribution (e.g. data). Although there are some arguments based on the concepts of "the wisdom of the crowd" trying to undermine or counter-balance the impacts of the individuals' biases on the collective decision, there are two challenges to this notion: Firstly, the representativeness, i.e. the structure of the crowd and "power of the elites" in many crowdsourcing projects have been questioned. This could be an issue in terms of biases, however some believe that the super active contributors are experts and so it is better to leave some decisions in their hands. While Antweiler et al. (2004), Giles (2005), Rajagopalan et al. (2011) and Shankland (2003) showed that collective decision-making can be more accurate than experts' comments, accuracy does not necessarily show all the aspects of quality and might not be even loosely correlated with potential bias. In terms of biases Greenstein et al. (2017) found the knowledge produced by the crowd are not necessarily less biased than the knowledge produced by experts. Nematzadeh et al. (2017) confirmed this by using Wikipedia contents, however, they found both biases and data quality could be moderated if substantial revisions and supervisions (of the gatekeepers) were implemented.

The second challenge to the notion of the "wisdom of the crowd", is the process of many VGI projects which is not based on collective decision but instead on crowd "participation". The difference is relatively implicit but highly important; the participants do not vote for/against every single decision or entry. The collection of individual decisions does not necessarily mean the collective decision making. Therefore the wisdom of the crowd may not be relevant to such projects as the individual bias can remain at micro-level. As the crowd makes decisions individually in a participatory project, the results of an individual's contributions could be biased. Therefore for these projects the case of "given enough eyeballs, all bugs are shallow" (Raymond 1998) is no longer valid as there is not enough revision/votes for each piece of information contributed by volunteers.

The following section focuses on the types of biases that may exist in VGI projects. Nematzadeh et al. (2017) found that crowd-sourced content can also produce a large sample with a great variety of biased opinions. Next subsection looks at the types and sources of biases that can exist in crowdsourced and volunteered geographic data.

2.1 Types and Sources of Bias

Biases can be categorised in many ways. Tripepi et al. (2010) categorises them as such: unmeasured confounders, selection bias (Heckman, 1990) and information bias (Hodgins et al., 1993).. In the context of this paper, selection bias can occur when 'wrong' contributors are selected/allowed to contribute. For example, if the residents of rural areas were selected to participate in a city transportation network related project. Due to the nature of VGI projects selection bias is one of the most important and influencing types of biases and also relatively hard to detect and treat.

While Tripepi et al. (2010) and Ricketts (1990) find it difficult to evaluate the impacts of selection bias, other studies that have investigated the impacts of selection bias and provided some solutions, including qualitative assessment, quantitative bias analysis and incorporation of bias parameters into the statistical analyses (Heckman, 1990; Munafò et al., 2017). From another perspective, an algorithm or method (including sampling method) can be categorised into ascertainment bias or systematic bias. Information bias, also known as observational bias, can occur when a type of error remains in a variable. For example GPS selective availability, if it had not been discontinued, could make GPS traces stored in the OSM database baised.

From another typology and classification perspective, VGI may suffer from the ascertainment bias, i.e. some members of the contributor population are less likely to contribute or be included. Ascertainment bias could seem to be related to the sampling and selection bias.

There are more than 300 types of bias and a wide range of classifications and this paper cannot and does not wish to review classifications of bias. While there are several classifications and clusters of sources of bias, this paper classifies the biases into two main classes of systematic and project level biases. Systematic biases exist in any crowdsourcing platform due to issues that will impact any project (e.g. population density or the impact of digital inequalities amongst age groups). The project level biases are due to the project nature, design and approaches (e.g. a culture that is not welcoming to people without high levels of education). Since the systematic biases are more common, we focus on them within trajectory data. This paper deliberately focuses on raw trajectory data, which is simply captured by the device with no further analysis of the contributors, and thus is likely to (a) reveal systematic biases, and (b) minimise the biases coming from the individual due to narrow thinking, shallow thinking, overconfidence, myopia and potential escalation of commitment (Soll et al., 2014). However this part provides a short list of some common/potential biases that can exist within VGI.

- ✓ Bandwagon bias (also referred to or related to groupthink bias and herd behavior) refers to the tendency of contributors to change their own opinion in favor or due to of an existing group pattern/behaviour to look they believe in the same way (Bikhchandani et al., 1992). This project can see this bias in the geography of the participants' movements. As shown in figure 1, participants tend to follow the same area/route even though they are allowed to go anywhere within the campus. Although this may be due to common Points of Interest (PoIs) to visit, such as restaurants and lecture theatres, it may also be influenced by bandwagon bias.
- ✓ Confirmation bias and congruence bias which can occur when a contributor looks for confirmatory patterns or interprets information in a way that confirms their assumptions and passes the hypothesis (Nickerson, 1998).
- ✓ Déformation professionnelle and Dunning-Kruger biases refer to the two extreme ends of level of expertise and knowledge which prevents a view of the world with the same eye as the crowd's (i.e. ignoring/forgetting the broader point of view) due to several reasons including overconfidence, getting used to the disciplines frameworks, illusion of superiority and lack of critical thinking (wu et al., 2018), (Friedman, 2017). These have been observed frequently in the attributes associated with the trajectory data that the participants contributed, i.e. over detailed information, relatively explanatory comments and in some cases overly analytical and interpretative opinions that only Geospatial Information Scientist would think of. Although this project tried to ignore the majority of the attributes associated with trajectory data, some of them need to be included or compared with inferred information (to assess the accuracy of the predictive analytics). They include speed of movement and travel mode, which are reported with too much detail. For example in some cases the participant measured the speed using three devices and reported the average of three. Also, some of the participants also tried to challenge the project by going to "hard to recognise" areas as they did know how the positioning system works. The exact opposite to this intentional deviation of the ordinary movement could be:
- ✓ Extreme aversion bias can occur when the participants would rather contribute in their "comfort zones" both metaphorically and geographically. This is basically due to the tendency of people to go to great lengths to avoid choosing an option that lays on the extremes of thinking (Madan et al., 2014).
- ✓ Framing bias leads to different conclusions from the same observations depending on how the observations are represented. This is very important for remote mapping projects where the base datasource could be represented in different ways. The tracking app was developed and deployed for both Android and iOS devices and so the difference in underlying maps and chipsets could influence the results.
- ✓ Selective reporting may occur when some features/events are more likely to be reported/contributed (Ioannidis et al., 2014). Many only report their activities or add a feature/attribute if they are thought to be interesting. This is particularly the case for social media check-ins and location sharing. In this project the participants tend to report their work-related travel way more than personal trips, even though they had expressed no objections to do with their privacy.
- ✓ Neglect of probability bias. This refers to the tendency to completely disregard the probability and uncertainty when making a decision under uncertainty (Fiedler et al., 2000).

- ✓ Status quo bias (related to loss aversion and endowment biases) refers to the tendency of contributors to focus on the same things (Samuelson et al., 1988), (Kahneman et al., 1991). Status quo bias could also result in a biased geographic distribution of PoIs.
- ✓ Lake Wobegon, self-serving and overconfidence bias is about reporting, believing and adding features about oneself and their properties in a self-promoting way (Kruger, 1991; Forsyth, 2008). For example adding features and self-promoting attributes/tags about their own properties. This is the opposite to modesty bias (Daniel et al., 2018).

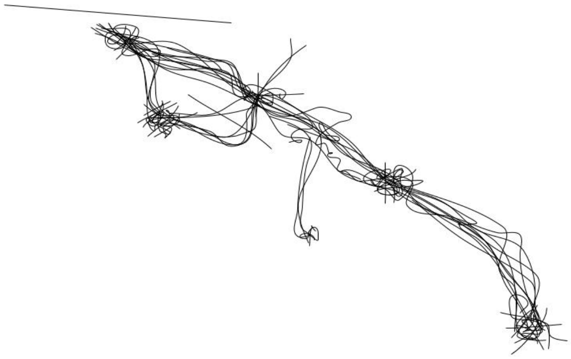

Figure 1. Trajectories of movements, consisting on 9982 segments (after noise reduction, outliers detection and Douglas-Peucker simplification and smoothing)

In addition to the cognitive biases that contributors might display, there are several technological biases that make the measurements as well as the equipment (such as positioning devices) biased. However, since this paper focuses on raw trajectory data, which is simply captured by the device with no further analysis of the contributor, these types of biases are not detectable and so may be better to assume that except the cognitive and social biases that might have an impact on the associated attributes and survey results, there are no further biases included in the raw trajectory data and their associated attributes.

3 Experiments and Results

This paper examines the impact of biases imposed on or already existing in the trajectory data on the output of spatio-temporal data mining process. The trajectories of movements, captured in Hanover over two months (July 2013 to August 2013) are used by another project (Basiri et al., 2016b) for automatic feature extraction and clusters recognition. This project examines the impact of biases on predicting the very same variables, i.e. the recognised features (PoIs) and travel mode.

To do so, the biases are gradually added to the data to see how the result of the very same data mining process would change. Each data mining process requires a training sample and control/test sample set. Input data are randomly divided into two sets and the patterns, rules, clusters etc are identified among the training sample (first set) and then they are used on the control data set to see how the results can predict the available data. However, in this paper, this random division has been intentionally ignored. The data are divided into two sets of training and control sets based on the identified sources of biases.

In order to intentionally make the training data biased, the input trajectories are divided into two sets of training and control datasets but not randomly (unlike the usual procedure of data mining). The two datasets are divided into training and control with respect to (a) time of the day, (b) day of the week, (c) weather conditions, (d) gender of contributor and (e) spatial and temporal density of trajectory in each 1kn grid square. These factors are identified by the Random Forest technique as the most important predictive variables. This means using a Random Forest method whereby the relative importance of each source of bias is identified. The parameters of weather conditions, length of the trajectory, time of day and gender are ranked with the highest importance (77.43%, 69.43%, 45.29%, and 11.4% respectively). The reason for applying Random Forest to recognise these factors is due to higher prediction accuracy (to predict travel mode and points of interests) in comparison with other Machine Learning methods (see figure 2). In fact, the impacts of biases are examined on the best performing technique for the available dataset, i.e. Random Forest.

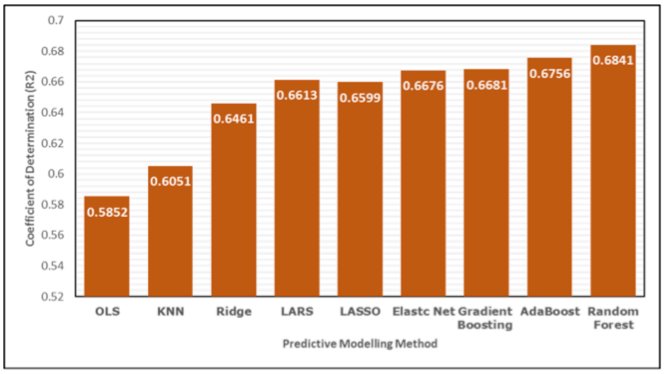

Figure 2. Prediction accuracy of various machine learning methods (elevation excluded)

Having training datasets selectively generated, the predictive models can be now trained and the models applied to predict the values of the control data. Then, firstly, the accuracy of the predictive model is compared against the accuracy of the predictive model that is based on random division (of training and control datasets). Secondly, the biases in the training dataset are gradually being moderated, i.e. balancing with supervision the skewness of the determining variable(s) distribution in both datasets. This will help to see (a) how the accuracy of predictive models are changing and (b) how the results would be different in terms of the nature of the micro/macro clusters of trajectories. Note that such gradual bias removal/moderation is different from holdout set or n-fold cross validation data split (Kohavi, 1995). As shown in figure 3, Cross-validation divides data into n subsets and then the model building and error estimation process is repeated n times.

While there is a best-practice of 30-70 percent split for holdout set or 1:n for cross-validation, as shown in figure 3, this paper divides the whole dataset into two (potentially equal but not necessarily with a pre-determined proportion) subsets with supervised bias imposed, and then gradually moderates the biases by continuously mixing the members of training and control sets. The important factor to divide the input trajectory data into two sets of training and control/test datasets is to have the maximum bias in terms of time of day, day of the week, weather conditions, gender and spatial and temporal density. The first round of data division could therefore have any ratio. The two sets will, of course, get moderated in the next rounds and could potentially increase/decrease in terms of size in order to make the datasets less and less skewed.

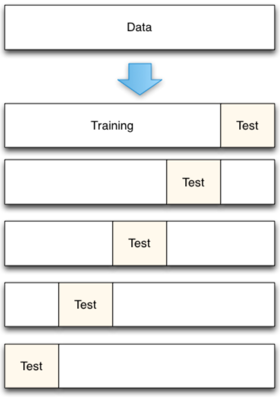

Figure 3. Stepwise fold (here 5-fold) cross validation (Kohavi, 1995)

Figure 4 shows the change in the accuracy of the predictive models with respect to the level of bias embedded in the training and control datasets. As shown in figure 4, the optimum level of accuracies for both the training and control data is 0.7001 (R2). This accuracy is surprisingly close to the level of accuracy of the Random Forest model where training and control data were generated randomly from the whole dataset.

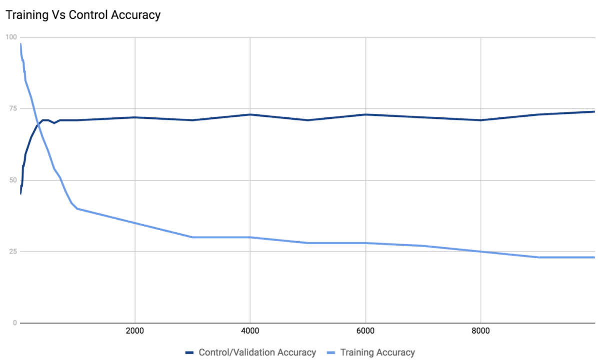

Figure 4. Accuracy of the predictive model for the training and control datasets, with respect to the rainy vs. cloudy or sunny weather conditions.

Predictably a predictive model which is fitted/trained based on a training dataset that only includes data belonging to a rainy day cannot predict the rest of data very accurately. However it doesn't necessarily mean that the accuracy for the control data is complementary, i.e. 1-(accuracy for extremely biased training set), as the variables are not independent and also the relationships might not be linear.

The overall accuracy of the predictive model for the training and control datasets could be optimised with variety of techniques, such as Bias-variance decomposition (Rodriguez et al., 2010; James et al., 1984). However for a relatively structured dataset, without a significant level of bias, the simplest approach could be division of the set randomly. As shown in figure 4, the optimum accuracy, i.e. where the two curves meet, is very close to the achieved accuracy of the Random Forest with random sampling (68.41% vs 70.01%).

Beside the overall accuracy, there is a potentially interesting outcome that seem good to discuss, however requires further studies. Using some clustering algorithms the trajectories could be clustered. A general clustering approach represents some trajectories with a feature vector, which denotes the similarity between trajectories by the distance between their feature vectors. This paper replicates the same approach used by Basiri et al (2016b), i.e. using the clustering of trajectory based on Hausdorff Distance (CTHD) metric to identify similarity with an adopted Micro- and Macro-clustering framework. In the CTHD, the similarity between trajectories is measured by their respective Hausdorff distances and they are clustered by the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm (Birant et al., 2007). The Hausdorff Distance metric is to identify similarity with an adopted Micro- and Macro-clustering framework.

The result of this experiment shows there are some micro-clusters, which are normally recognised as anomalies in less-biased data, and may have some meaningful patterns or features. However, they will be rejected/ignored by control datasets as a valid microcluster by very high level of confidence. This is mainly due to the fact that such micro-cluster only exist as minority and it would be unlikely to have them in the control dataset too. So when very biased data are used, the number of micro-clusters increases and the number of macro-clusters decreases. However, in the control mode, only macro clusters remains. The confidence for micro-clusters is relatively low while the confidence for the macro-cluster is even higher than standard data mining process. It seems the less biased data increase the confidence in general, although they may miss some micro-clusters. Figure 5 shows some of the biggest micro/micro clusters/PoIs, a few of which still ignored by the less/de-biased dataset.

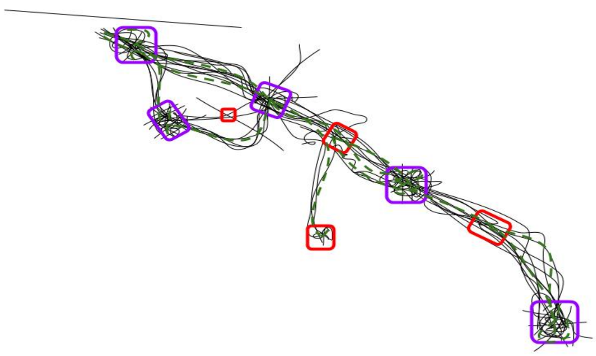

Figure 5. Green lines: Macro clusters trajectories, Purple: valid/confirmed PoIs, Red: disapproved or rejected PoIs

Conclusion

Crowdsourced and voluntarily gathered data have brought an unprecedented opportunity to researchers to analyse and extract knowledge, particularly using statistical and machine learning techniques. However, all VGI and crowdsourcing projects are in biased some way, most obviously due to voluntary response samples rather than random sampling. There are also other types and sources of bias within the crowdsourced datasets. This project studies the impacts of biases on the results of ML algorithm and evaluates how the input biased data can influence the results, in terms of accuracy and reliability, of the learnt patterns. To do so, the biases in the raw trajectory data are intentionally added and then the accuracy of the training and control data for several ML techniques are measured. This paper found the random selection of the training and control datasets result in the level of accuracy that is very close to the accuracy that optimises both training and control models while biases are intentionally being imposed to the Random Forest training and control datasets. In addition, the results of trajectory mining from very biased data showed the recognition of some micro-clusters, which are normally recognised as anomalies in less-biased data that may have some meaningful patterns or features. However, they are not visible in the control datasets, as such micro-clusters may only exist as minorities to also have them identified in the control dataset. So less biased data increases the confidence in general, although some micro-clusters maybe missed.

References

Basiri, Anahid, Mike Jackson, Pouria Amirian, Amir Pourabdollah, Monika Sester, Adam Winstanley, Terry Moore, and Lijuan Zhang. 2016a "Quality assessment of OpenStreetMap data using trajectory mining." International Journal of Geospatial information science 19, no. 1 (2016): 56-68.

Basiri, A., Amirian, P., & Mooney, P. 2016b. Using crowdsourced trajectories for automated OSM data entry approach. Sensors, 16(9), 1510.

Bell, S., Cornford, D. and Bastin, L., 2015. How good are citizen weather stations? Addressing a biased opinion. Weather, 70(3), pp.75-84.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. "A theory of fads, fashion, custom, and cultural change as informational cascades." Journal of political Economy 100, no. 5 (1992): 992-1026.

Birant, Derya, and Alp Kut. "ST-DBSCAN: An algorithm for clustering spatial–temporal data." Data & Knowledge Engineering 60, no. 1 (2007): 208-221.

Brown, G., 2017. A review of sampling effects and response bias in internet participatory mapping (PPGIS/PGIS/VGI). Transactions in GIS, 21(1), pp.39-56.

Daniel, Florian, Pavel Kucherbaev, Cinzia Cappiello, Boualem Benatallah, and Mohammad Allahbakhsh. "Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions." ACM Computing Surveys (CSUR) 51, no. 1 (2018): 7.

DeMaio, Theresa J. "Refusals: Who, where and why." Public Opinion Quarterly 44, no. 2 (1980): 223-233.

Di Zhu, Zhou Huang, Li Shi, Lun Wu & Yu Liu (2017) Inferring spatial interaction patterns from sequential snapshots of spatial distributions, International Journal of Geographical Information Science, 32:4, 783-805, DOI: 10.1080/13658816.2017.1413192

Fiedler, Klaus, Babette Brinkmann, Tilmann Betsch, and Beate Wild. "A sampling approach to biases in conditional probability judgments: Beyond base rate neglect and statistical format." Journal of Experimental Psychology: General 129, no. 3 (2000): 399.

Forsyth, Donelson R. "Self-serving bias." (2008): 429.

Friedman, Hershey H. "Cognitive Biases that Interfere with Critical Thinking and Scientific Reasoning: A Course Module." (2017).

Gardner, Z., Mooney, P., Dowthwaite, L. & Foody, G., 2018. Gender differences in OSM activity, editing and tagging. Proceedings of GISRUK 2018 Conference, Leicester, 17-20th April.

Gardner, Z. & Mooney, P., 2018. Investigating gender differences in OpenStreetMap activities in Malawi: a small case-study. Proceedings of AGILE Conference, Lund, Sweden, 12-15th June.

Giles, J. 2005. "Internet Encyclopaedias Go Head to Head," Nature (438:7070), pp. 900–901.

Goyder, John. "Surveys on surveys: Limitations and potentialities." Public Opinion Quarterly 50, no. 1 (1986): 27-41.

Greenstein, Shane, and Feng Zhu. "Do Experts or Crowd-Based Models Produce More Bias? Evidence from Encyclopædia Britannica and Wikipedia." Forthcoming, Management Information Systems Quarterly (2017).

Haklay, M., Singleton, A. and Parker, C., 2008. Web mapping 2.0: The neogeography of the GeoWeb. Geography Compass, 2(6), pp.2011-2039.

Hammond, Kenneth R. "Coherence and correspondence theories in judgment and decision making." Judgment and decision making: An interdisciplinary reader (2000): 53-65

Heckman, James. "Varieties of selection bias." The American Economic Review 80, no. 2 (1990): 313-318.

Hinds, P.J., 1999. The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. Journal of Experimental Psychology: Applied, 5(2), p.205.

Hodgins, Holley S., and Miron Zuckerman. "Beyond selecting information: Biases in spontaneous questions and resultant conclusions." Journal of Experimental Social Psychology 29, no. 5 (1993): 387-407.

Howard, A., Zhang, C. and Horvitz, E., 2017, March. Addressing bias in machine learning algorithms: A pilot study on emotion recognition for intelligent systems. In Advanced Robotics and its Social Impacts (ARSO), 2017 IEEE Workshop on (pp. 1-7). IEEE.

Ioannidis, John PA, Marcus R. Munafo, Paolo Fusar-Poli, Brian A. Nosek, and Sean P. David. "Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention." Trends in cognitive sciences 18, no. 5 (2014): 235-241.

James, Lawrence R., Robert G. Demaree, and Gerrit Wolf. "Estimating within-group interrater reliability with and without response bias." Journal of applied psychology 69, no. 1 (1984): 85.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. "Anomalies: The endowment effect, loss aversion, and status quo bias." Journal of Economic perspectives 5, no. 1 (1991): 193-206.

Kohavi, Ron. "A study of cross-validation and bootstrap for accuracy estimation and model selection." In Ijcai, vol. 14, no. 2, pp. 1137-1145. 1995.

Kruger, Justin. "Lake Wobegon be gone! The" below-average effect" and the egocentric nature of comparative ability judgments." Journal of personality and social psychology 77, no. 2 (1999): 221.

Madan, Christopher R., Elliot A. Ludvig, and Marcia L. Spetch. "Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions." Psychonomic bulletin & review 21, no. 3 (2014): 629-636.

McKenzie, Craig RM. "Rational models as theories–not standards–of behavior." Trends in Cognitive Sciences 7.9 (2003): 403-406.

Mooney, P. and Corcoran, P., 2014. Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. Transactions in GIS, 18(5), pp.633-659.

Mooney, R.J., 1996. Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning. arXiv preprint cmp-lg/9612001.

Mullen, W.F., Jackson, S.P., Croitoru, A., Crooks, A., Stefanidis, A. and Agouris, P., 2015. Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features. GeoJournal, 80(4), pp.587-605.

Munafò, Marcus R., Kate Tilling, Amy E. Taylor, David M. Evans, and George Davey Smith. "Collider scope: when selection bias can substantially influence observed associations." International journal of epidemiology (2017).

Neis, P., Goetz, M. and Zipf, A., 2012. Towards automatic vandalism detection in OpenStreetMap. ISPRS International Journal of Geo-Information, 1(3), pp.315-332.

Nematzadeh, A., Ciampaglia, G.L., Menczer, F. and Flammini, A., 2017. How algorithmic popularity bias hinders or promotes quality. arXiv preprint arXiv:1707.00574.

Nickerson, Raymond S. "Confirmation bias: A ubiquitous phenomenon in many guises." Review of general psychology2, no. 2 (1998): 175.

Rajagopalan, M. S., Khanna, V. K., Leiter, Y., Stott, M., Showalter, T. N., Dicker, A. P., and Lawrence, Y. R. 2011. "Patient-oriented Cancer Information on the Internet: A Comparison of Wikipedia and a Professionally Maintained Database," Journal of Oncology Practice (7:5), pp. 319–323.

Raymond, E. 1998. "The Cathedral and the Bazaar," First Monday, http://tinyurl.com/bqfy3s, accessed May 2018

Ricketts, John A. "Powers-of-ten information biases." MIS Quarterly (1990): 63-77

Rodriguez, Juan D., Aritz Perez, and Jose A. Lozano. "Sensitivity analysis of k-fold cross validation in prediction error estimation." IEEE transactions on pattern analysis and machine intelligence 32, no. 3 (2010): 569-575.

Salk, C., Sturn, T., See, L. and Fritz, S., 2016. Local knowledge and professional background have a minimal impact on volunteer citizen science performance in a land-cover classification task. Remote Sensing, 8(9), p.774.

Samuelson, William, and Richard Zeckhauser. "Status quo bias in decision making." Journal of risk and uncertainty 1, no. 1 (1988): 7-59.

Soll, J., Milkman, K. and Payne, J., 2014. A user's guide to debiasing.

Tripepi, Giovanni, Kitty J. Jager, Friedo W. Dekker, and Carmine Zoccali. "Selection bias and information bias in clinical research." Nephron Clinical Practice 115, no. 2 (2010): c94-c99.

Wu, Kaidi, and David Dunning. "Hypocognition: Making sense of the landscape beyond one's conceptual reach." Review of General Psychology 22, no. 1 (2018): 25.