A framework for investigating illegal wildlife trade on social media with machine learning

Di Minin, Enrico

2019-02


http://hdl.handle.net/10138/298334
https://doi.org/10.1111/cobi.13104

Downloaded from Helda, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
Diversity

A framework for investigating illegal wildlife trade on social media with machine learning

Enrico Di Minin, Christoph Fink, Tuomo Hiippala, and Henrikki Tenkanen

1 Digital Geography Lab, Department of Geosciences and Geography, University of Helsinki, Helsinki, FI 00014, Finland
2 Helsinki Institute of Sustainability Science, University of Helsinki, Helsinki, FI 00014, Finland
3 School of Life Sciences, University of KwaZulu-Natal, KwaZulu-Natal, Durban 4000, South Africa
4 Department of Languages, University of Helsinki, Helsinki, FI 00014, Finland

Unsustainable harvesting is one of the major threats driving the global extinction crisis (Maxwell et al. 2016). Wildlife trade is a multibillion-dollar industry, in which thousands of animals, plants, and associated products are traded globally as food, pets, medicines, clothing, and trophies (Dalberg Global Development Advisors 2012). Wildlife trade escalates into a crisis when an increasing proportion is illegal and unsustainable and thus directly threatens the persistence of many species in the wild (Ripple et al. 2016). High-profile species, such as rhinoceroses and elephants (Wittmer et al. 2014; Di Minin et al. 2015a), as well as many lesser-known species (Rosen & Smith 2010; Phelps & Webb 2015), are threatened by illegal trade. Illegal wildlife trade is among the largest illegitimate businesses (Dalberg Global Development Advisors 2012). Furthered by poverty, poorly monitored borders, corruption, and weak regulations and enforcement, illegal wildlife trade continues to grow (Dalberg Global Development Advisors 2012; UNODC 2016).

In recent years, the scale and nature of illegal wildlife trade has changed dramatically, and the internet has become a major market for wildlife products (Lavorgna 2014). High-profile species, such as rhinoceroses and elephants (Wittmer et al. 2014; Di Minin et al. 2015a), as well as many lesser-known species (Rosen & Smith 2010; Phelps & Webb 2015), are threatened by illegal trade. Illegal wildlife trade is among the largest illegitimate businesses (Dalberg Global Development Advisors 2012). Furthered by poverty, poorly monitored borders, corruption, and weak regulations and enforcement, illegal wildlife trade continues to grow (Dalberg Global Development Advisors 2012; UNODC 2016).

We propose a new research framework in which machine learning is used to investigate illegal wildlife trade on social media platforms (Fig. 1). The framework has 3 stages: mining, filtering, and identifying relevant data on illegal wildlife trade on social media.

User-generated content, including images, text, and videos, can be downloaded from several social media platforms, including Facebook, Twitter, Weibo, and Flickr, via an application programing interface (API) (see https://www.programmableweb.com/category/social/apis?category=20087 for a full list). Application programing interfaces are publicly available, and researchers can independently collect global-scale data from the content made available by the social media company. Using APIs, researchers can access publically available data. Social media data collected via APIs are being used increasingly in conservation (e.g., Di Minin et al. 2015b), but automated classification is limited. Automated content classification can help filter out information irrelevant to illegal wildlife trade (e.g., “pangolin armoured vehicle” as opposed to pangolin taxa [Fig. 1a]) and render content classification cost-efficient.

Machine learning and its subfields and components (deep learning, neural networks, and natural language processing) can be used to identify verbal, visual, and audiovisual content pertaining to illegal wildlife trade.

*email enrico.di.minin@helsinki.fi

Article impact statement: Machine learning can be used to automatically monitor and assess illegal wildlife trade on social media platforms. The above article from Conservation Biology, published online on 12 March 2018 in Wiley Online Library (wileyonlinelibrary.com) has been revised to eliminate overlap with a previously published article: Machine Learning for Tracking Illegal Wildlife Trade on Social Media in Nature Ecology and Evolution 2, 406–407 (2018). Paper submitted October 30, 2017; revised manuscript accepted February 23, 2018.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Conservation Biology. Volume 33, No. 1, 210–213
© 2018 The Authors. Conservation Biology published by Wiley Periodicals, Inc. on behalf of Society for Conservation Biology.
DOI: 10.1111/cobi.13104
Figure 1. Framework to (a) mine, (b) filter, and (c) identify relevant data on the illegal wildlife trade from social media platforms with machine learning. Photo in (c) is from Twitter.
another application, not from social media, Hernandez-Castro and Roberts (2015) developed an automated system to detect potentially illegal elephant ivory items for sale on eBay.

Although the characteristics of social media data provide a great opportunity to track illegal wildlife trade, there are still challenges and caveats (e.g., spatial inaccuracy and unreliable data related to scams, etc.) associated with using social media content for research purposes (Di Minin et al. 2015b; Tsou 2017). In addition, scientists and practitioners have the ethical responsibility to minimize potential harm to people who share illegal wildlife trade content on social media platforms (Zook et al. 2017). For example, the privacy policy and terms of use of each social media platform should be followed strictly and only publicly available social media data used. The anonymity of social media users should be respected and their privacy protected by anonymizing the data so that it cannot be linked to any personal information, such as names or phone numbers. Receiving, storing, processing, and applications of social media data should strictly follow all data security and privacy requirements (e.g., the European Union General Data Protection Regulation) of the country where researchers are based. Another problem is that a wealth of relevant data on illegal wildlife trade is currently not open to research via APIs. For this reason, manual observation, filtering and classification of content, particularly to assess whether content pertains to legal or illegal trade, remains important (Hinsley et al. 2016; Eid & Handal 2017). Still, openly available data, which can be downloaded through the APIs, represent an important sample of all available social media data. Our framework can be applied within the safe environments of social media platforms without breaching privacy fences.

Our proposed methods and analyses are relevant for the implementation of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (e.g., Decisions 17.92 and 17.93 at the Conference of the Parties 17). Given the pressing issue, creating partnerships between CITES parties, social media companies, and scientists working on artificial intelligence will help create the conditions (e.g., by accessing full social media data in full respect of privacy) that will make investigation of the illegal wildlife trade on social media possible. Our framework, with differences in related to how data can be downloaded, can also be applied to other online platforms.

Acknowledgments

An earlier version of this paper was published online in Accepted Articles. It was removed and substantially revised to eliminate overlap with an earlier publication by the authors on the same topic. The authors acknowledge an error in judgment that resulted in the need for this revision. E.D.M thanks the Academy of Finland 2016–2019 (grant 296524) for support. C.F. thanks the University of Helsinki for support via an Early Career Grant to E.D.M.

H.T. was funded by the Finnish Cultural Foundation. H.T. thanks the DENVI doctoral program at University of Helsinki for support.

Literature Cited


