Sentiment Analysis: ready for conservation

Margreet Drijfhout, Dave Kendal, Dany Vohl, and Peter T Green

We welcome the paper by Ladle et al. (2016), as it brings attention to the important topic of understanding social responses and cultural trends through the analysis of digital texts. The availability of these data, along with the number of tools created to analyze them, has increased immensely with the arrival of, and growing access to, the internet. We agree with Ladle et al. (2016) that conservation scientists have only just begun to tap into this wealth of data. One topic they mention briefly is the analysis of sentiment, which is described as "particularly difficult to automatically assess, owing to the semantic complexity of language and the frequent use of abbreviations, misnomers, misspellings, slang, and sarcasm" (Ladle et al. 2016). We believe the field of Sentiment Analysis has progressed far beyond its infancy and is worthy of much greater attention in conservation. Sentiment Analysis offers a more powerful and nuanced way of analyzing digital text, which can provide a more mechanistic understanding of the way people think about conservation issues, beyond what is possible by simply tracking changes in word frequencies.

Sentiment Analysis measures people's opinions, sentiments, evaluations, attitudes, and emotions from written language (Liu 2012). Initially used in computer science for analyzing sentiments expressed in product reviews, the use of Sentiment Analysis has been rapidly spreading to other scientific disciplines (Figure 1; Liu 2012). It has the potential to harvest relevant information from an ever-growing volume of social media data (Hirschberg and Manning 2015). For instance, sentiment expressed in tweets has been used to predict not only (1) the results of public opinion polls, highlighting the potential of tweets as a supplement or substitute for expensive and time-intensive traditional polling (O'Connor et al. 2010), but also (2) damage from a natural disaster (Kryvasheyeu et al. 2016) and trading opportunities in financial markets (Zheludev et al. 2014).

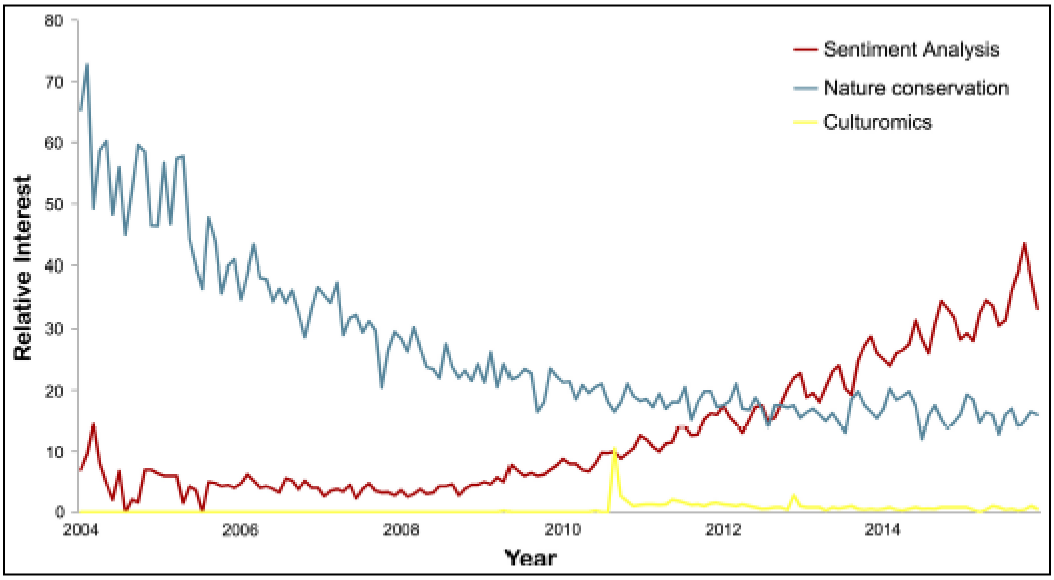

Figure 1. Google Trends analysis of the terms "nature conservation", "sentiment analysis", and culturomics. The relative frequency of a search term for every month over the past 12 years is shown.

At the core of Sentiment Analysis is a lexicon containing words or phrases expressing different sentiments (Liu 2012). Lexicons can be manually created as a list of predefined words or automatically created using machine learning techniques (eg training datasets) (Yu and Hatzivassiloglou 2003). Sentiment Analysis can identify different "opinion dimensions" such as polarity (positive, negative, or neutral), strength (the intensity of positive or negative sentiments), and emotion (categories including sadness, joy, etc) (Bravo-Marquez et al. 2014). It is true that language has complexities, like sarcasm or irony, which makes sentiment difficult to accurately characterize (Ladle et al. 2016). However, other complexities – such as acronyms, slang, and misspelled words – are being tackled by the increasing sophistication of software (Bravo-Marquez et al. 2014), which is leading to Sentiment Analysis accuracy rates as high as 90% (Yu and Hatzivassiloglou 2003).

Today, there is great interest in the analysis of social media data for conservation purposes, for instance to measure public engagement with conservation, or to understand preferences for biodiversity (Di Minin et al. 2015). Sentiment Analysis of opinion-based data could provide important insights into public opinion on sensitive or controversial conservation issues. For example, public outcry over culling native animals can influence conservation policy and management, yet there is usually little empirical evidence of public opinion beyond that expressed by a few involved stakeholders. Sentiment could also be analyzed over space to monitor changes in public response to specific parks, or over time to measure responses to particular management actions. This approach has clearly progressed beyond its infancy, and should be seen as a valuable analytical tool in the rapidly expanding world of social media data analysis for conservation purposes.

Bravo-Marquez F, Mendoza M, and Poblete B. 2014. Meta-level sentiment models for big social data analysis. Knowl-Based Syst 69: 86–99.

Di Minin E, Tenkanen H, and Toivonen T. 2015. Prospects and challenges for social media data in conservation science. Front Environ Sci 3: 63. Hirschberg J and Manning CD. 2015. Advances in natural language processing. Science 349: 261–66.

Kryvasheyeu Y, Chen H, Obradovich N, et al. 2016. Rapid assessment of disaster damage using social media activity. Sci Adv 2: e1500779.

Ladle RJ, Correia RA, Do Y, et al. 2016. Conservation culturomics. Front Ecol Environ 14: 269–75. Liu B. 2012. Sentiment analysis and opinion mining. San Rafael, CA: Morgan & Claypool Publishers.

O'Connor B, Balasubramanyan R, Routledge BR, and Smith NA. 2010. From tweets to polls: linking text sentiment to public opinion time series. Proceedings of the 4th International AAAI Conference on Weblogs and Social Media; 23–26 May 2010; Washington, DC. Menlo Park, CA: AAAI Press.

Yu H and Hatzivassiloglou V. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing; 11–12 Jul 2003; Sapporo, Japan. Stroudsburg, PA: Association for Computational Linguistics.

Zheludev I, Smith R, and Aste T. 2014. When can social media lead financial markets? Sci Rep 4: 4213.