Tell me a story about the birds and the bees: Using NLG to foster public engagement in nature conservation projects

Advaith Siddharthan
University of Aberdeen
advaith@abdn.ac.uk

Chris Mellish
University of Aberdeen
c.mellish@abdn.ac.uk

Hien Nguyen
University of Aberdeen
n.nguyen@abdn.ac.uk

Cheng Zeng
University of Aberdeen
c.zeng@abdn.ac.uk

Imogen Pearce
University of Aberdeen
i.pearce@abdn.ac.uk

René van der Wal
University of Aberdeen
r.vanderwal@abdn.ac.uk

ABSTRACT

For many important nature conservation programmes, western societies are increasingly reliant on the activities of volunteers, who, collectively, have come to represent an unpaid work force of considerable size and importance. Although a variety of effective ways exist to generate adequate recruitment, volunteer retention is harder to achieve, particularly when schemes grow bigger or tasks get more difficult. We describe two case studies that we are using to investigate the hypothesis that richness of information provision, of the kind that can be provided by Natural Language Generation (NLG), can play a role in fostering volunteer interest and motivation. Both these case studies involve collaboration with large existing conservation projects, which provide the possibility for evaluation on a realistic scale.

1. DIGITAL TOOLS & CITIZEN SCIENCE

There is increasing realisation of the potential of digital approaches, including the use of websites and social media, to increase participation in "citizen science", which includes observing and monitoring the natural world. For instance, in the UK, the Open Air Laboratories (OPAL) network (www.opalexplorenature.org) is a large current initiative led by Imperial College, which aims to create and inspire a new generation of nature-lovers by getting people to explore, enjoy and protect their local environment [3]. Within OPAL, iSpot (www.ispot.org.uk) is an online nature community that connects beginners with experts and fellow enthusiasts. Other groups have investigated the use of standard social networking sites to generate public interest and collect data about the distribution of species [4]. Publicly available resources include www.thewildlab.org, which provides software for a number of mobile platforms, and www.scienceforcitizens.net, which acts as a forum for citizen scientists to find out about projects they can participate in and for researchers to publicise their projects.

Although digital tools can enthuse the public and be used to enlist (for a short time at least) willing volunteers for nature conservation projects, initiatives such as the above have to contend with at least the following issues:

1. Data quality: participants are generally untrained, and not necessarily motivated to produce high-quality data [4]. This is not a problem if a project is primarily an "outreach" activity, but many projects also have scientific goals.
2. Retention of volunteers: in order to secure continuing participation, systems need to be constantly new and interesting, and engage with users.

Digital Engagement '11, November 15–17, 2011, Newcastle, UK.

2. CASE STUDIES

2.1 Public engagement and red kites

Narratives that bring to life the behaviour of specific members of a species can have a positive effect on the involvement of citizens in conservation activities involving that species.

We are investigating the above hypothesis by building an NLG system that is able to produce a commentary on the travels of red kites that are tagged with satellite tracking devices. The red kite is a species reintroduced to the UK, whose success varies in currently unexplained ways across regions. For this project, we are working with one of the largest nature conservation charities in Europe.

Our aim in the red kite project is to bring these tagged birds of prey "to life" by constructing narratives around their daily activities. Currently, the satellite tag data is being used by the charity to manually create blogs such as:

> "...Ruby (Carrbridge) had an interesting flight down to Loch Duntelchaig via Dochfour on the 6th March before flying back to the Drumsmittal area, spending the 10th March in the Loch Ussie area (possibly also attracted by the feeding potential there!) and then back to Drumsmittal for the 13th..."

Our interest, in addition to automating the generation of such blogs (following earlier data-to-text work such as [2]), is in making these narratives more interesting, by using the data to illustrate key aspects of red kite behaviour. In this regard, this project requires ecologists and computer scientists to work together closely. As an illustration, we provide two different views of GPS fixes from one tagged red kite. Fig. 1 (top) shows how far the kite travels from its nest over time. This can be used to construct narratives around the exploratory behaviour of red kites in their early years. Fig. 1 (bottom) shows the same GPS data, but in a spatial view, plotting latitude against longitude without regard to time. This analysis highlights the kite's favoured locations as well as its broader range, providing another view of the bird's behaviour. Clearly such data can also be superimposed onto GIS layers based on weather, habit or wider land use data to help construct more informative narratives.

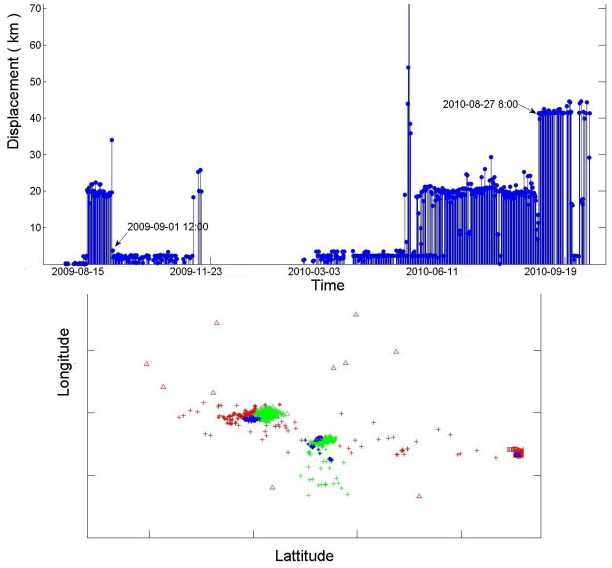

Figure 1: Plot of (top) distance from nest as a function of time and (bottom) clusters of locations visiting for one tagged Red Kite.

2.2 Volunteer recordings of bumblebees

Volunteers can learn identification skills better if they receive tailored feedback than if they receive fixed feedback.

Citizen science projects aim to collect vast quantities of data from untrained volunteers. In the ecological domain, the most successful of these ventures is the RSPB's big garden birdwatch (http://www.rspb.org.uk/birdwatch), which attracted over 600,000 volunteer recorders this year. Given that the common garden birds in the UK are easily identifiable, such volunteer collected data can give an accurate picture of trends for some species. However, for less common species, or species that are harder to spot or identify, such citizen science efforts are less reliable and also attract and retain fewer volunteers [1].

In this project we are attempting to use NLG to increase volunteer motivation and volunteer accuracy in identifying different species of bumblebees. There are 25 different species of bumblebee in the UK; many of these are threatened by habitat loss. This is a matter of concern as bumblebees are important pollinators of wild and garden flowers as well as crops and fruit trees. Bumblebee identification requires some training and, particularly when identifying from photographs, can be quite challenging. Thus, both volunteer motivation and accuracy are key issues for a citizen science project in this domain. The bumblebee charity that we work with runs a scheme where volunteers submit photos of bumblebees through a web interface, which are then identified by an expert. This approach doesn't scale and the question is whether we can train volunteers to accurately identify the bees, by giving tailored feedback in this process. For this purpose, we have designed an interface (Fig. 2) that allows volunteers to try and identify a bumblebee from a photo. There is an identification guide, where users can select features that are visible in the photo (colour patterns on thorax and abdomen, shape of head and tail, colour of wings, etc.) to prune the list of possibilities. We have two aims:

1. To provide useful feedback when users make an incorrect identification, to improve identification skills.

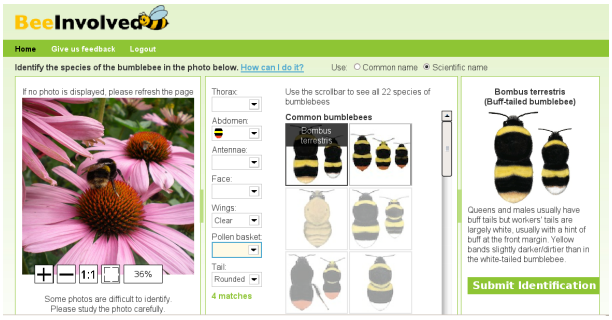

Figure 2: Bumblebee identification interface.

2. To use "crowd-sourcing" techniques to reduce the effort required from the domain expert.

For 1., we are recording the user's mouse events as they select bumblebee features and consider different species. For instance, in Fig. 2, the user has clicked on *Bombus terrestris* in the middle panel to see details on the right, suggesting that this was an option that was considered. Further, the user has selected features such as thorax, tail shape and wing colour, indicating that only the rear of the bee is visible in the photo. Now, if a user has misclassified a picture, we can try to understand the user's choice based on features in common with the correct choice. We can also explain why the classification is incorrect by highlighting features that exist only for the correct bee. This form of feedback can be further personalised based on the species considered by the user, the orientation of the bee in the photo and the user's past identification record and motivation levels.

For 2., the key questions are whether we can use multiple users to increase the accuracy of volunteer recordings, and can we train volunteers sufficiently that they provide reliable classification for the vast majority of photos? We have run experiments which show that agreement between untrained volunteers correlates well with the accuracy of the agreed classification. We are currently running experiments to test accuracy and return rates over time for users provided with different levels of feedback.

3. CONCLUSIONS

This paper introduces a new project applying Natural Language Generation to conservation initiatives based around public engagement. We are applying NLG both to communicate data from satellite tagged animals, and to train and motivate volunteers in ecological recording programmes.

4. REFERENCES

[1] McCaffrey, R. Using citizen science in urban bird studies. *Urban Habitats 3*, 1 (2005), 70–86.

[2] Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., and Sykes, C. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence 173*, 7-8 (2009), 789–816.

[3] Silvertown, J. A new dawn for citizen science. *Trends in Ecology & Evolution 24*, 9 (2009), 467–471.

[4] Stafford, R., Hart, A., Collins, L., Kirkhope, C., Williams, R., Rees, S., Lloyd, J., and Goodenough, A. Eu-Social Science: The Role of Internet Social Networks in the Collection of Bee Biodiversity Data. *PloS one 5*, 12 (2010), e14381.