Blogging birds: Generating narratives about reintroduced species to promote public engagement

Advaith Siddharthan, Matthew Green, Kees van Deemter, Chris Mellish & René van der Wal
{advaith, mjgreen, k.vdeemter, c.mellish, r.vanderwal}@abdn.ac.uk
University of Aberdeen

Abstract

This paper proposes the use of NLG to enhance public engagement during the course of species reintroductions. We examine whether ecological insights can be effectively communicated through blogs about satellite-tagged individuals, and whether such blogs can help create a positive perception of the species in readers' minds, a requirement for successful reintroduction. We then discuss the implications for NLG systems that generate blogs from satellite-tag data.

1   Introduction

Conservation of wildlife is an objective to which considerable effort is devoted by governments and NGOs across the world. A variety of web-based approaches can help make the natural world more accessible to the public, which in turn may translate into greater public support for nature conservation initiatives. The present paper explores the role of Natural Language Generation (NLG) in bringing up-to-date information about wild animals in their natural environment to nature enthusiasts.

We focus on the reintroduction of the red kite to the UK. This member of the raptor family was once widespread in the UK, but prolonged and intense persecution led to its near extinction. Since 1989, efforts have been ongoing to reintroduce this species in various locations across the country. We are working together with one of the largest nature conservation charities in Europe to use NLG for public engagement around a small number of satellite-tagged reintroduced red kites.

The public engagement activities surrounding this reintroduction initiative have two subtly different objectives: (i) to communicate ecological insights to increase awareness about the species, and (ii) to create a positive image of the reintroduced species to harness public support for the reintroduction. Currently, data from these satellite tags are 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...

Such blogs are used by schools which have adopted individual kites, and pupils can read these texts alongside a map plotting the GPS locations of 'their' kite. As can already be seen from the above, there is currently little ecological information about the species in these blogs. Because of the perceived importance of education to the success of reintroductions, there is a clear desire to include more ecological insights. Yet, time and resource limitations have prevented the charity from doing so; they perceive the writing of such blogs already as very time consuming, and indeed, rather mundane.

In this paper, we explore the use of blogs based on satellite tag data for communicating ecological insights and creating a positive image of a species. We consider both aspects, deemed essential for a successful species reintroduction, and focus on how the blogs can be made more informative than those currently being written by the charity.

2   Related work

Data-to-text systems (e.g., Goldberg et al. (1994); Theune et al. (2001); Portet et al. (2009)) have typ-
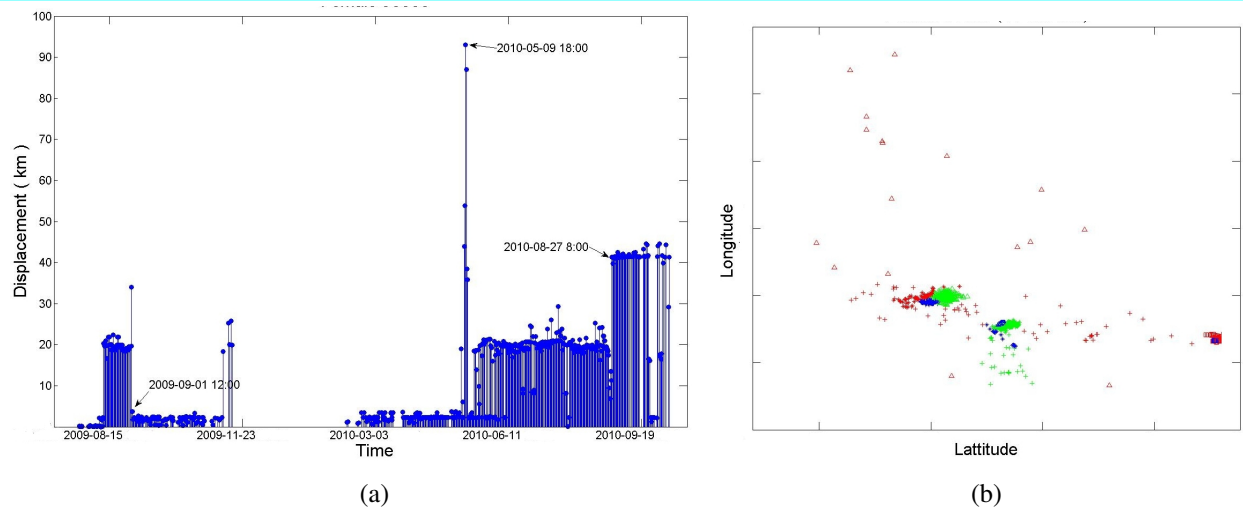
120

Figure 1: Plot of (a) distance from nest as a function of time, and (b) clusters of visited locations.

ically been used to generate summaries of technical data for professionals, such as engineers, nurses and oil rig workers. There is some work on the use of data-to-text for lay audiences; e.g., generating narratives from sensor data for automotive (Reddington et al., 2011) and environmental (Molina et al., 2011) applications, generating personal narratives to help children with complex communication needs (Black et al., 2010), and summarising neonatal intensive care data for parents (Mahamood et al., 2008).

Our application differs from the above-mentioned data-to-text applications, in that we aim to generate inspiring as well as informative texts. It bears some resemblance to NLG systems that offer "infotainment", such as Dial Your Disc (Van Deemter and Odijk, 1997) and Ilex (O'Donnell et al., 2001). In fact, Dial Your Disc, which generates spoken monologues about classical music, focused emphatically on generating engaging texts, and achieved linguistic variation through the use of recursive, syntactically structured templates (see also, Theune et al. (2001)). We intend to extend a data-to-text system in similar ways, using ecological insights to make narratives engaging for non-experts.

3 Overall Goals

Our overall aim is to bring satellite tagged animals (in this case study, red kites) "to life" by constructing narratives around their patterns of movement. We require individual locations of a bird to be explained in the context of its wider spatial use, and the ecological interpretations thereof. This paper has the following goals:

1. To illustrate how satellite tag data can be analysed to identify behavioural patterns for use in generating blogs (content selection);

2. To test whether blogs written by an ecologist based on such data analysis can be used to educate as well as create a positive perception of the species;

3. To investigate the challenges for NLG in automating the generation of such blogs.

4 Data analysis for identifying behaviours

From an NLG perspective, our interest in automating the generation of blogs from satellite tag data is in making these narratives more interesting, by using the data to illustrate key aspects of red kite behaviour. To illustrate how we can relate the data to behaviours, we provide two graphical views of GPS fixes from a tagged red kite. Fig. 1(a) shows how far a focal kite is located from its nest over the course of a year. We propose to use such data to construct narratives around ecological insights regarding the exploratory behaviours of red kites during their first year after fledgling. Fig. 1(b) shows the same GPS data, but now spatially, thereby plotting latitude against longitude of all fixes without regard to time. This portrayal highlights the kite's favoured locations (indicated in different colours based on a MATLAB cluster analysis which automatically estimates the parameters of a Gaussian mixture model,

even when clusters overlap substantially), as well as its broad range.

These plots illustrate two key aspects of kite behaviour: *exploration* and *site-fidelity* (the presence of favoured locations that the kite tends to return to). In addition, we are interested in communicating various *feeding behaviours* as well as that, unlike many other birds of prey, red kites are *social* birds, often found in groups. Feeding and social behaviours cannot be directly identified from the data. However, they can often be inferred; for instance, a red kite spending its time by the side of a main road is likely to be looking to scavenge on road kill.

5   Study on engaging readers using blogs

We now report a study that explores whether such ecological insights can be effectively communicated through blogs constructed around an individual of the species, and whether such blogs can help create a positive perception of the species in a reader's mind.

This study was based on a text manually constructed by an ecologist based on five weeks of data such as in Fig 1 from a red kite named "Red Baroness". For this study, the data was mapped onto a simplified world with seven features: a lake, a shoreline, fields, a road, a moor, a forest and a river. A sample of the text is shown in Figure 2 for illustration.

Week 2: How different the pattern of movements of Red Baroness was this week! On Monday, she shot off past Bleak Moor, on her longest journey so far north-east of the lake. She appeared not to find much of interest there, and on the next day she was observed combing the edges of Green Park, possibly in search of a group of birds resting in the top half of the trees. The bird was clearly restless however, as on Thursday she was observed following River Rapid, downstream for further than she had been last month, finally stopping when she reached Blue Lake again.

Figure 2: Sample material showing week 2 from the five week blog

5.1   Experimental Design

80 participants were shown the material: a five week blog on the movements of the focal red kite, named Red Baroness, alongside a picture of a red kite and a schematic map marking the seven features of interest. Participants were students at the University of Aberdeen. The experiment was conducted in a lab in a supervised setting. After reading and returning the blog, each participant was asked to (a) summarise the blog they had just read in 5 lines, (b) state what they found most interesting, and (c) state what they did not like about the blog. These textual responses were manually coded for whether the four behaviour types (site fidelity, exploration, feeding and social) were identified by each participant.

To gauge the participants' perceptions of the kite, we used two methods. First, we asked the participant to answer four questions that tested various aspects of their willingness to engage with red kite conservation:

Q1   Would you be willing to contribute money to a charity that tries to protect kites?

Q2   The use of rat poison also leads to the death of kites that feed on the bodies of these rats. Would you be willing to sign a campaign against rat poison?

Q3   Should governments allocate more money than they do currently to protect kites from extinction?

Q4   Write your email if you wish to be sent more blogs.

Further to this, participants were asked to assess the red kite's personality. We follow (Gosling et al., 2003), who use the 44 question Big Five Inventory (BFI) (John et al., 1991; John et al., 2008) to assess the personality of dogs. We are interested in whether readers did assign personalities to the red kite in the blog and, if so, what these personality profiles looked like.

5.2   Results

We now analyse the extent to which our participants were informed about red kite ecology as well as how willing they were to engage with conservation efforts and how they perceived the species.

5.2.1   Informativeness

More than half the participants identified feeding behaviour (61%) and social (54%) behaviour. The other two ecological concepts were not mentioned explicitly in the blog that participants read, but needed to be inferred. Around a quarter of participants managed to infer the notion of site fidelity

(23%), the most difficult of the concepts, and 41% inferred exploratory behaviour.

5.2.2 Engagement

39% provided their email address to receive further blogs (the only real commitment), and an equal number expressed willingness to contribute money for red kite conservation efforts. 85% expressed willingness to sign a campaign against rat poisoning, and 61% wanted increased government spending for red kite conservation.

We detected a correlation between recall/inference of behaviours and willingness to engage (plotting total number of behaviours recalled/inferred by each participant against the total number of engagement questions answered affirmatively, $r_{pearson} = 0.31$; $p < 0.005$; $n = 80$). One interpretation of this result is that greater insights into the life of this bird has positively influenced the reader's perceptions of it. Further qualitative studies are needed to substantiate this, but we view this result as evidence in favour of incorporating ecological insights into the blogs.

5.2.3 Perception

Table 1 shows the big five personality traits assigned to Red Baroness by participants. The BFI is constructed such that being non-committal about the 44 trait questions would result in scores of 3. The ability of readers to assign human personality traits (significantly different from 3.0) to the red kite indicates a willingness to anthromorphise the bird. The last column shows the average personality of 21 year old humans (from Srivastava et al. (2003)), which is the same age group as our participants. The values for extroversion, agreeableness and conscientiousness are very similar, and the kite has lower neuroticism and openness.

6 Implications for NLG

The above study indicates that it is possible to use narratives based on satellite tag data to communicate ecological insights as well as create a positive perception of the species in the readers' minds. To generate texts that are fluent and engaging enough that readers will be both informed and entertained by them poses challenges that are sharply different from the ones facing most data-to-text systems,

Trait	Red Kite	Conf. Int.	21 yo
Extroversion	3.28	3.07–3.48	3.25
Agreeableness	3.64	3.47–3.80	3.64
Conscientiousness	3.48	3.26–3.69	3.45
Neuroticism	2.60	2.41–2.80	3.32
Openness	3.29	3.11–3.47	3.92

Table 1: Big five personality traits of Red Baroness with 99.9% confidence intervals, compared to average 21 year olds (6076 people) (Srivastava et al., 2003)

whose primary purpose is to offer decision support. Our goals are more similar to those of Dial Your Disc (Van Deemter and Odijk, 1997), with the added requirement that texts should be easy to read. For instance, ecological concepts (such as site fidelity) could be communicated by explicitly defining them. However, we would prefer these to be inferred from more engaging narratives.

The blogs currently created by the charity (cf. Section 1) are, stripped down to their essence, a sequence of locations. We propose to interlay these sequences of locations with descriptions of red kite behaviours, broadly categorised as fidelity, exploration, feeding or social. Algorithm 1 outlines the planning process. We have developed an initial prototype that implements this for our simplified world. Using template based generation, we can automatically generate blogs such as the following for arbitrary sequences of locations in our simplified world:

> This week Red Baroness continued to feel like stretching her wings. On Monday she was seen in the fields by the lake, calling out to other kites. On Tuesday and Wednesday she stayed along the road, looking for roadkill on the country lanes. On Thursday she returned to the fields by the lake – clearly there was plenty to eat there.

To scale this up to the real world, work is in progress to augment our data analysis component by using a variety of GIS data to map geo-coordinates to habitat, terrain and demographic features from which we can identify relevant kite behaviours.

Our remaining challenges are to (a) compile a large list of red kite behaviours, (b) use paraphrasing approaches to create variety in descriptions of behaviour and (c) develop means to interweave more

1. Identify place names of interest to the user among the many GIS locations frequented by the red kite

2. For each place of interest (ordered by time):

   (a) describe place in terms of relevant geographical features

   (b) describe one or two behaviours (feeding or social) associated with any of these features

   (c) make a reference to any exploratory behaviour or site fidelity if identified from previous sequence.

Algorithm 1: Generate a blog about a red kite

complex behaviours, such as mating, into the narratives. There is ongoing interdisciplinary work into each of the above. Variation is likely to be critical to the endeavour as these blogs are aimed at engaging the reader, not just at presenting information. This can be achieved both by expanding the range of behaviours we describe, and the range of ways we can realise these through language.

7 Conclusions

This paper reports a study that informs the application of NLG technologies to conservation efforts centred around public engagement. We report on findings which indicate that it is possible to use narratives loosely based on satellite tag data to communicate ecological insights as well as to create a positive perception of the species in readers' minds. This informs an approach to automating the creation of blogs from satellite-tagged red kites by interleaving sequences of locations with descriptions of behaviour. A proof of concept system has been developed for a simplified world, and is in the process of being scaled up to the real world, using GIS data.

Acknowledgments

References

R. Black, J. Reddington, E. Reiter, N. Tintarev, and A. Waller. 2010. Using nlg and sensors to support personal narrative for children with complex communication needs. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pages 1–9. Association for Computational Linguistics.

E. Goldberg, N. Driedger, and R.I. Kittredge. 1994. Using natural-language processing to produce weather forecasts. *IEEE Expert*, 9(2):45–53.

S.D. Gosling, V.S.Y. Kwan, and O.P. John. 2003. A dog's got personality: a cross-species comparative approach to personality judgments in dogs and humans. *Journal of Personality and Social Psychology*, 85(6):1161.

O.P. John, E.M. Donahue, and R.L. Kentle. 1991. The big five inventoryversions 4a and 54. *Berkeley: University of California, Berkeley, Institute of Personality and Social Research*.

O.P. John, L.P. Naumann, and C.J. Soto. 2008. Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, pages 114–158.

S. Mahamood, E. Reiter, and C. Mellish. 2008. Neonatal intensive care information for parents an affective approach. In *Computer-Based Medical Systems, 2008. CBMS'08. 21st IEEE International Symposium on*, pages 461–463. IEEE.

M. Molina, A. Stent, and E. Parodi. 2011. Generating automated news to explain the meaning of sensor data. *Advances in Intelligent Data Analysis X*, pages 282–293.

M. O'Donnell, C. Mellish, J. Oberlander, and A. Knott. 2001. Ilex: an architecture for a dynamic hypertext generation system. *Natural Language Engineering*, 7(3):225–250.

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

J. Reddington, E. Reiter, N. Tintarev, R. Black, and A. Waller. 2011. "Hands Busy, Eyes Busy": Generating Stories from Sensor Data for Automotive applications. In *Proceedings of IUI Workshop on Multimodal Interfaces for Automotive Applications*.

S. Srivastava, O.P. John, S.D. Gosling, and J. Potter. 2003. Development of personality in early and middle adulthood: Set like plaster or persistent change?. *Journal of Personality and Social Psychology*, 84(5):1041.

M. Theune, E. Klabbers, J.R. de Pijper, E. Krahmer, and J. Odijk. 2001. From data to speech: a general approach. *Natural Language Engineering*, 7(01):47–86.

K. Van Deemter and J. Odijk. 1997. Context modeling and the generation of spoken discourse. *Speech Communication*, 21(1-2):101–121.