Human Computation and Crowdsourcing: Works in Progress and Demonstration Abstracts
An Adjunct to the Proceedings of the Second AAAI Conference on Human Computation and Crowdsourcing

Groupsourcing: Problem Solving, Social Learning and Knowledge Discovery on Social Networks

Jon Chamberlain
School of Computer Science and Electronic Engineering
University of Essex
Wivenhoe Park, Colchester CO4 3SQ, England
jchamb@essex.ac.uk

Abstract

Increasingly social networks are being used for citizen science, where members of the public contribute knowledge to scientific endeavours. Tasks can be presented and solved using human computation, termed groupsourcing, with users benefiting from community tuition and experts gaining knowledge from the crowd. This paper gives details of a prototype that utilises groupsourcing to solve image classification tasks, to support social learning and to facilitate knowledge discovery in the domain of marine biology.

Introduction

Citizen science, where members of the public contribute knowledge to scientific endeavours, is an established research methodology and social networks have also been successfully used to connect professional scientists with amateur enthusiasts (Sidlauskas et al. 2011). Social networks are self-organised and decentralised; tasks are created by the users, so they are motivated to participate, and the natural language of the interface allows them to express their emotions, appreciation and frustrations whilst solving the tasks.

Groupsourcing, where a task is completed using a group of intrinsically motivated people of varying expertise connected through a social network, is an effective method of human computation in some domains (Chamberlain 2014). Users solve problems with high accuracy, educate each other and share novel information and ideas. Contribution to science, learning and discovery are the driving motivations behind citizen science participation (Raddick et al. 2013).

There are drawbacks with this method of crowdsourcing: data is unstructured, not archived and difficult to analyse using standard natural language processing techniques; access is not simple; and there is an uneven distribution of workload and content creation.

This paper gives details of a prototype application[1] that utilises groupsourcing to solve image classification tasks in the domain of marine biology and shows how aggregation and visualization techniques can support social learning and knowledge discovery.

[1] http://www.jonchamberlain.com/groupsourcing

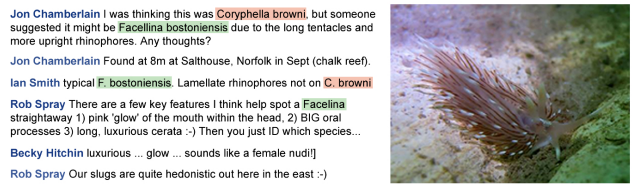

Figure 1: Detail of a typical message containing an image classification task having been analysed for named entities.

Structuring Data

Social networks such as Facebook, LinkedIn and Flickr offer access to large user communities through integrated software applications and/or a back-end API. Data held on Facebook can be accessed using the Graph API[2] and temporarily cached for processing, a procedure described in detail elsewhere (Chamberlain 2014).

Ontologies, gazetteers or controlled vocabularies can be used to structure the content, although not without problems. This prototype uses, in the first instance, an ontology of marine species[3] as a hierarchical list of named entities. Each chunk of text from a message thread is scanned for named entities from the ontology and an index table is created in a MySQL database. The prototype temporarily stores a cache of the messages however a larger implementation would only need the index table to be stored, with message content stored on the social network. Query expansion is used for spelling variations, for example it is common for marine species names to be abbreviated by concatenating the genus such as "Coryphella browni" to "C. browni" - see Figure 1.

The named entity index is used to aggregate the messages across groups allowing a user to find all content regarding a particular marine species and other species that are associated with it (what it eats, what it looks similar to, etc.).

Additionally, messages containing a named entity with an image attached are used to create a gallery of photographic examples of the species - see Figure 2. Only links to the images are stored, the images themselves are hosted on the social network. Each image is credited with the author's name.

[2] https://developers.facebook.com/docs/graph-api

[3] http://www.marinespecies.org (Sept 2012)

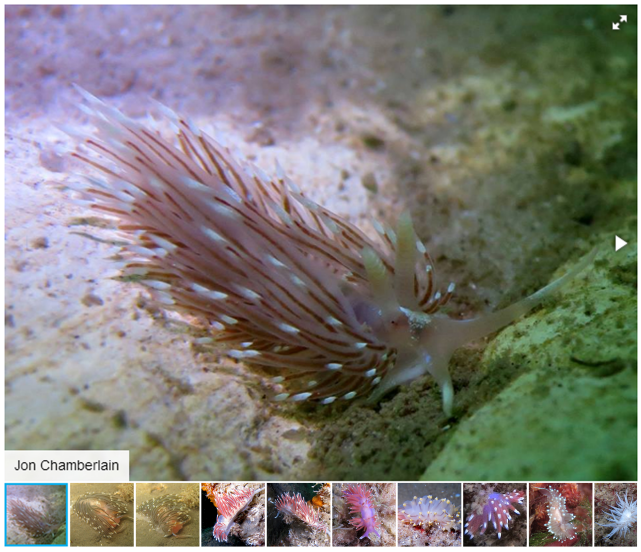

Figure 2: Screenshot of the aggregated image gallery.

Challenges

The natural language processing of the message threads is a significant challenge and needs to cope with ill-formed grammar and spelling, contextual referencing and sentiment, for example:

> "Is this Coryphella browni or bostoniensis?"
> "I don't think this is C. brownii."
> "I agree with you on that."

Additionally, the ontology and identifying morphology of marine species is in constant flux, meaning identifications previously considered correct may have changed. For example, there was a significant update to the taxonomic group Chromodorididae (Johnson and Gosliner 2012) that rendered many static Web resources and books out of date, however users frequently correct identifications to the new nomenclature on social networks.

The aggregated image set for a named entity uses simple methods to order the images, with the most credible appearing first, however inappropriate images will need to be removed with some form of user filtering.

One of the most serious drawbacks is the changing technology and popularity of social media. Although reasonably mature with a high take-up rate, it is still an emerging technology, and changes are made to the terms of service, access and software language that could swiftly render a dependent system redundant.

Social Learning

Outreach and communicating knowledge to the general public is a core objective of academic institutions and social networks can be used to facilitate these aims. Social learning, where users on the social network teach and support each other in an ad-hoc manner, encourages users to engage in the learning process to an extent that suits their interests and time restraints.

This prototype compiles the knowledge of social network groups to be used as a resource for supporting social learning. Some users will learn enough to be able to answer other user's questions reducing the traditional bottleneck of a few experts having to do the majority of the work.

Knowledge Discovery

The opportunity to discover something unknown is a driving user motivation behind citizen science. Image classification projects that use crowds to tag unknown objects are popular and can lead to significant scientific discoveries (Clery 2011). The enthusiasm of the public to participate was most recently seen with the search for missing Malaysia Airlines flight MH370 in 2014 where millions of users analysed satellite imagery, tagging anything that looked like wreckage, life rafts and oil slicks.[4]

Using human computation to correct machine output, with the most interesting cases being forwarded to experts, could be the most efficient and cost effective way of using the wisdom of the crowd (Eickhoff 2014).

This prototype gives expert users access to important information in message threads. In the case of marine biology these are associated species, temporal shifts, geographic range and other niche dimensions that could be indicators of ecosystem changes caused by pollution, overfishing or climate change.

Future Work

Future work will focus on scalability and the temporal nature of the data, including the message stream and ontologies. Further processing will be applied to extract more information from the thread dialogue in order to develop more sophisticated tools, such as matrix keys, facet searching and niche models.

References

Chamberlain, J. 2014. Groupsourcing: Distributed problem solving using social networks. In *Proceedings of HCOMP14*.

Clery, D. 2011. Galaxy evolution. Galaxy Zoo volunteers share pain and glory of research. *Science* 333(6039):173–5.

Eickhoff, C. 2014. Crowd-powered experts: Helping surgeons interpret breast cancer images. In *Proceedings of GamifIR'14*.

Johnson, R. F., and Gosliner, T. M. 2012. Traditional taxonomic groupings mask evolutionary history: A molecular phylogeny and new classification of the chromodorid nudibranchs. *PLoS ONE* 7(4):e33479.

Raddick, M. J.; Bracey, G.; Gay, P. L.; Lintott, C. J.; Cardamone, C.; Murray, P.; Schawinski, K.; Szalay, A. S.; and Vandenberg, J. 2013. Galaxy Zoo: Motivations of citizen scientists. *ArXiv e-prints*.

Sidlauskas, B.; Bernard, C.; Bloom, D.; Bronaugh, W.; Clementson, M.; and Vari, R. P. 2011. Ichthyologists hooked on Facebook. *Science* 332(6029):537.

[4] http://www.tomnod.com/nod/challenge/mh370_indian_ocean