Domain Analysis

Currently in place

MammalWeb's purpose is to collect information about wildlife in the UK and allow biological scientists to form conclusions about the ecosystem on both the national and local scale. This information can then be used to inform decisions about conservation projects, culling action and similar activities that affect wildlife and the environment.

The front end of MammalWeb at the moment is the website MammalWeb.org on which users can upload images from camera traps that they have set up in their local area. Anyone can then sign on as a 'spotter' and classify images drawn from the all those that have been uploaded. The classifications (species present and age/gender/number if relevant) are then stored in a database.

Currently, extracting useful information from these classifications is not well implememted - the data dump we recieved contained data from a little over 20000 photos with nearly 90000 individual classifications and there is no automated system in place for removing 'bad' information and condensing the useful data into a form that is easier to use for the biologists. Conclusions that have so far been drawn from the collected data has been done manually.

We have been provided with an algorithm designed to aggregate data collected in citizen scientist projects as outlined in Swanson et al. (2015). To summarize, the algorithm classifies a photo as being whatever the most common species that has been classified is. For example, if 8 classification of deer, 3 of horse and 1 of nothing here have been recorded then the Swanson algorithm will say that the photo is of a deer. 3 metrics to show how likely the classification is to be correct are then calculated-

1. Eveness: All non-blank classifications are used in Pielou’s eveness index to calculate this. The formula is where S is the number of different species classified as being present (2 in the example given) and pi is the proportion of total classifications for species i (8/11 and 3/11 in the example- remember that blanks are ignored here). This comes out as 0.845 for the example. If only one species is classified, the result will be 0 and the highest possible result is 1 so a relatively high result like 0.845 can be interpreted as high uncertainty that the aggregate classification is wrong.
2. Fraction blanks: The fraction of “nothing here” classifications for an image that has an aggregate classification that is not “nothing”. The above example would have fractional blanks of 1/12
3. Fraction support: The fraction of classifications that support the aggregate answer (in the above example it would be 2/3)

Related systems

A closely related system would be Snapshot Serengeti where a very similar system of uploading and classifying camera trap photos by citizen scientists is in place.

Zooniverse is another similar system, although it is not as closely related to Mammalweb as Snapshot Serengeti. Zooniverse is a massive compendium of varied scientific projects from wildlife surveying very similar to MammalWeb/Snapshot Serengeti to astronomical research on distant galaxies and research on weather and climate. The similarity with MammalWeb comes from the use of citizen scientists to process data on scales far beyond what can be efficiently realised with a small number of trained scientists.