Information Sciences 306 (2015) 21–33

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Using aggregation functions to model human judgements of species diversity

Gleb Beliakov [a], Simon James [a,*], Dale G. Nimmo [b]

[a] Deakin University, School of Information Technology, 221 Burwood Hwy, Burwood, Victoria 3125, Australia
[b] Deakin University, School of Life and Environmental Sciences, 221 Burwood Hwy, Burwood, Victoria 3125, Australia

ARTICLE INFO

Article history:
Received 9 October 2014
Received in revised form 1 February 2015
Accepted 6 February 2015
Available online 13 February 2015

Keywords:
Aggregation functions
Weights learning
Bonferroni mean
Species diversity
Ecological indices

ABSTRACT

In environmental ecology, diversity indices attempt to capture both the number of species in a community and the relative abundance of each. Many indices have been proposed for quantifying diversity, often based on calculations of dominance, equity and entropy from other research fields. Here we use linear fitting techniques to investigate the use of aggregation functions, both for evaluating the relative biodiversity of different ecological communities, and for understanding human tendencies when making intuitive diversity comparisons. The dataset we use was obtained from an online exercise where individuals were asked to compare hypothetical communities in terms of diversity and importance for conservation.

1. Introduction

When policy makers and governments talk about the importance of preserving biodiversity, they have an intuitive rather than mathematical idea of what diversity is. Models that inform management decisions and ecology research, however, require notions like species diversity to be quantified. The debate surrounding which indices should be used is often argued with reference to some key properties and simple examples that show one or another index to be counter-intuitive, however there is no real basis upon which one can be deemed as more acceptable than another, so decision makers are able to choose the index that best serves their agenda. Although communities with no diversity (only one species present) and perfect diversity (a high number of species all equally abundant) can be easily defined, comparisons in diversity toward the middle of the spectrum are more difficult to formally articulate, and to date, there has been no study of whether existing models generally reflect the human perception of what it means for a forest or grassland area to be high or low in biodiversity. In this article, we apply the theory of aggregation functions in order to uncover the implicit judgements that are made when humans evaluate species diversity. We provide a method for making diversity assessments based on samples of elicited comparison judgements from experts, and use a real data set to show that the resulting evaluations correspond better with intuitive human perceptions than some of the popularly used diversity indices.

Aggregation functions have been successfully employed across a broad range of human decision-making contexts to uncover the influence and importance of variables and predict human judgements, e.g. in journal ranking [2], group decision-making [23,12] and image analysis [16].

* Corresponding author. Tel.: +61 3 9251 7481.
E-mail addresses: gleb@deakin.edu.au.au (G. Beliakov), sjames@deakin.edu.au (S. James), dale.nimmo@deakin.edu.au (D.G. Nimmo).

http://dx.doi.org/10.1016/j.ins.2015.02.013
0020-0255/

The real data used was gathered from an online exercise we conducted where participants were asked to intuitively judge whether one community was more abundant than another. To inform their decision, they were provided only with the information that is currently used to evaluate species diversity, namely, the individual abundances for all species present in the community.[1] From these supplied pairwise comparisons, we can see how well the indices used in ecology literature reflect these judgements, and further, whether results from the theory of aggregation functions and weight learning could potentially provide biodiversity evaluations that are closer to our human perceptions. Many of the evaluations for evenness and diversity that have been developed and used in ecology can themselves be considered in the framework of aggregation functions.

We found that from the simple comparison data obtained from our online survey, aggregation functions could be constructed that were more consistent with the human evaluations than any one index alone. These functions aggregated the following indices:

- **total abundance** – the number of individuals present or observed across all species in the community;
- **richness** – the number of different species-types;
- **evenness** – an index reflecting how evenly the different species are distributed, which is usually calculated from richness and the individual abundances;
- **species diversity** – an index that usually is said to incorporate both richness and evenness, commonly calculated as their product.

Even when we only include total abundance, richness and evenness in the aggregation, we can still produce better results than the existing species diversity indices in terms of modeling the human assessments. In particular, we found that weighted Bonferroni means had the best overall performance, which can be explained by their ability to model mandatory requirements, i.e. they can model semantics such as:

IF *richness* is high **and** the average of *evenness* and *abundance* is high, THEN the community is high in biodiversity.

The advantage of using parameter-learning over an axiomatic approach is that we can model the subtle intuitive trade-offs that are made by experts without having to specify them *a priori*. In quantifying biodiversity, there is clearly a trade-off that occurs between having a high number of species present with some in low abundance, and having a lower number of species whose abundances are all equal. Although the aggregation functions used can still be interpreted in terms of reasonable properties and behavior, determining their parameters according to expert opinions allows a further justification in that they conform to human expectations and intuitions.

The article will be set out as follows. In Section 2, we present the preliminary concepts required for conducting our analysis. To this end we provide the background on aggregation functions and how to learn their weight parameters using linear programming. In Section 3, we give an overview of ecological indices used for quantifying species diversity and provide details of the online exercise we conducted and the dataset obtained. We then show how aggregation functions can be used to predict human judgements of species diversity in Section 4, and then show the analysis side of this process, i.e. how aggregation functions can be used to interpret the tendencies of the respondents, in Section 5. In Section 6 we provide some discussion and indicate some of our future research directions and finally Section 7 concludes.

2. Preliminaries

We firstly set out the necessary background in aggregation functions which we will use to formulate evaluations of species diversity. After giving some of the important definitions, properties and families, we provide an overview of how we can fit aggregation functions to data along with some recent results and techniques.

2.1. Aggregation functions

Aggregation functions lie at the heart of many decision processes where it is necessary to summarize an input set with a single value. Although many of the results have arisen independently in various research fields, overviews of construction methods and properties with a particular focus on soft computing and decision making can be found in [9,14,21].

We will use the following definition.

Definition 1. An aggregation function $f : [0,1]^n \to [0,1]$ is a function non-decreasing in each argument and satisfying $f(0,\dots,0) = 0$ and $f(1,\dots,1) = 1$.

Aggregation functions can be defined for inputs other than those given over the unit hypercube, however for the moment we will restrict our considerations to this case. Scaling is employed where necessary in our fitting procedures to ensure that all inputs are given over the same interval.

Depending on the context, the aggregation function desired may belong to a specific class.

[1] We make the special note that our perception of diversity may be dependent on other factors, such as *which* species, how rare they are, how different they are to one another, and so on, however none of the current indices take these factors into account.