Information Sciences 258 (2014) 312–327

Contents lists available at ScienceDirect

Information Sciences

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

A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science

Gloria Bordogna [a,b,*], Paola Carrara [b,1], Laura Criscuolo [b,1], Monica Pepe [b,1], Anna Rampini [b,1]

[a] CNR-IDPA, via Bassini 9, 20133 Milano, Italy
[b] CNR-IREA, via Bassini 15, 20133 Milano, Italy

ARTICLE INFO

Article history:
Available online 1 August 2013

Keywords:
Citizen science
Volunteered geographic information
Information quality assessment
Soft constraint
Fuzzy aggregation operator

ABSTRACT

The paper analyses the challenges and problems posed by the use of Volunteered Geographic Information (VGI) in citizen science and a proposal is formulated for assessing VGI quality based on a linguistic decision making approach so as to allow its feasible use for scientific purposes.

VGI quality is represented by indicators at distinct levels of granularity which take into account the distinct components of the VGI items. The quality indicators represent both the extrinsic quality, depending on the characteristics and reputation of the sources of information; the intrinsic quality, depending on the distinct accuracy and precision of information; and, last but not least, the pragmatic quality, depending on the user needs and intended purposes. In order to assess the pragmatic quality of VGI items, a linguistic decision making approach is defined that allows users to rank and finally filter the VGI items based on the satisfaction of distinct criteria expressed by means of both linguistic terms, defining soft constraints on the distinct quality indicators, and linguistic aggregators, defining fuzzy operators which combine the satisfaction degrees of the soft constraints at distinct hierarchical levels to yield the final satisfaction of the VGI items. Finally, an example of quality assessment in a glaciological citizen science project is discussed.

1. Introduction

"Citizens Science" has been coined to indicate projects whose activity cannot be carried out solely by the scientific team but needs the help of a large number of volunteer citizens to carry out some of the tasks [23].

In most of the citizen science projects volunteers are asked to provide information of various forms and of various nature, such as textual notes, pictures, measures about objects of interest. One common dimensions of the information is the geographic one, also known as Volunteer Geographic Information (VGI) [19]: a VGI item is the unit of information that describes an entity of interest that either has a location on Earth, or has been observed in a given location. For example, in tourist applications entities of interests may be museums, restaurants, hotels, and a VGI item is known as Point Of Interest (POI). In the glaciers' monitoring application we will discuss in this paper, a VGI item is a unit of information comprising pictures created by a volunteer to allow glaciologists monitoring the temporal variation of glacier's extent [11].

The contribution of volunteers to science is not new. In museums there are hundreds of millions of plants and animals which have been mostly collected by volunteers over the years. What changes with citizen's science is the sharp increase

* Corresponding author. Visiting reseacher at CNR-IREA, via Bassini 15, 20133 Milano, Italy. Tel: +39 02 236993295.
E-mail addresses: gloria.bordogna@idpa.cnr.it (G. Bordogna), carrara.p@irea.cnr.it (P. Carrara), criscuolo.l@irea.cnr.it (L. Criscuolo), pepe.m@irea.cnr.it (M. Pepe).
[1] Tel.: +39 02 236993456.

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.07.013

in both the quantity, and timely of VGI, that makes it possible to use it for monitoring processes thus constituting a new challenge for science.

Nevertheless, many researchers are critical to the usability of VGI due to its heterogeneous quality. In order to use VGI for science it is necessary to analyse the possible causes of its quality degradation and define or adopt policies for both regulating and constraining the data entry by volunteers so as both minimizing the possible creation of weak VGI items characterized by errors, ambiguities, incompleteness and inaccuracy and defining mechanisms for allowing users to assess the desired quality of VGI items depending on the purpose of their research activity.

In fact VGI quality assessment implies the consideration of several factors, both extrinsic with respect to the content of VGI items, depending on the characteristics and reputation of the sources of information, i.e., of the volunteers, which are difficult to quantify precisely; intrinsic factors, depending on VGI contents which may have accuracy and precision different for the textual notes, the images, the measurements, etc.; and, last but not least, pragmatic factors which depend on the intended purposes of the VGI within the project and thus can vary both from project to project and within the same project based on user needs and expectations.

For example, the *eBird* project (http://ebird.org/content/ebird/) strongly stressed the need for *high* accuracy of the species classification of the birds observed by volunteers across North America, and stated that "*a database is only as good as its weakest record*". Thus, if even a single VGI item is found as not accurate enough, *all* VGI items created by the same volunteer are deemed questionable.

In the glaciological project [11] the objective is to retain only the VGI items that provide *neat* pictures of the glaciers taken *close* to historical observation stations.

Besides differing according to the goal of the project the criteria for the filtering of the VGI items are often expressed linguistically by terms such as *very neat*, *close,* whose meaning is subjective. Then, one needs to apply a decision making process that is capable to cope with the subjectivity of the selection criteria, expressed linguistically by the decision maker, to first rank the alternatives, i.e., the VGI items, and then to filter the ones that satisfy the criteria to a minimal acceptance level.

This is the reason that motivates our proposal to define a linguistic multi criteria decision making process to allow project coordinators, more generally named users, to flexibly filter the VGI items on the basis of multiple distinct criteria expressed linguistically. This is modelled with fuzzy set theory by calling upon the approaches defined in fuzzy decision making and in fuzzy databases for evaluating flexible queries [6,9,18,40].

To this end, VGI quality is first represented by indicators with distinct granularity and by taking into account the distinct components, of VGI items. Each VGI item is then associated with metadata, which contain the quality indicators. Finally the metadata are made available to the user for flexible querying.

In this paper, we first review the distinct types of citizen science projects and the expectations on the quality of the information they deal with; in Section 2. we summarise the main approaches to deal with the quality of VGI in citizen science projects; in Section 3. we propose distinct quality indicators for the VGI components, and in Section 4. the quality assessment is defined as a linguistic decision making process evaluated by a flexible query with linguistic terms to the database. Finally we discuss an application example in the context of a glaciological citizen science project [11].

2. Citizen science projects and information quality

Relying on volunteer citizens for scientific purposes has become feasible with the diffusion of the multimedia Web and Web 2 and the ease of acquiring and publishing on the Internet geo-referenced free texts and data, images and video, also by means of crowd-sourcing applications such as Google Earth, Flickr, Amazon Mechanical Turk, etc. The interest for such topic is testified by the growing number for scientific initiatives that aim at collecting and organizing experiences of volunteers that are eager to aid a scientific project.[2]

In order to understand the characteristics of the quality of volunteers' information and the scientists' expectations on the quality of the contributions we have analysed the citizen science projects from distinct points of views [5] .

One first crucial aspect that impacts on both information quality and scientists expectations, besides the *application field* which determines significant choices in data format, referring standards, and vocabulary, is related to the need of VGI.

Strictly speaking VGI is intended as information relative to geographic objects like the one collected in both the *Open Street Map* (http://www.openstreetmap.org/), the *Open Geo-data* (OGD) and the *OpenAddresses* (OA) projects [32], the aim of which is to collect geo-coded addresses by volunteers in a central database, and the *Wikimapia* project (http://wikimapia.org/) whose objective is to provide a crowd sourced geographic gazetteer, with the resulting datasets available to all at no charge. These geographical citizen science projects need geographically homogeneously distributed VGI with high geographic accuracy.

Nevertheless, VGI in a wide sense comprises also *projects in which the objects of interest are not geographic ones*, but their geographic reference (geographic footprint) is relevant information such as in most projects about objects identification (birds, insects, plants and so on). The projects do not need very accurate VGI of the area where they saw the object, but ask for accurately interpreting the meaning of each observation, like in the eBird project (http://ebird.org/content/ebird/) where high accuracy is needed regarding the species of the observed bird.

[2] see for instance"*Anyone Can Be A Citizen Scientist*"(http://www.adventureandscience.org/) , and "science Biddies" (http://www.sciencebuddies.org/).