A new look at semantic accuracy

Ola Ahlqvist
Department of Geography
The Ohio State University
Columbus, OH, USA
Ahlqvist.1@osu.edu

Fifteen years ago, Francois Salge provided a seminal introduction to the idea of semantic accuracy (Salge, 1995). In this presentation I will not revisit his work but rather try to provide a fresh look at semantic aspects of geographic information from a diverse, but necessarily incomplete, set of perspectives. I begin with some observations from a recent study, The Synchronous Objects project (http://synchronousobjects.osu.edu), developed at the Ohio State University's Advanced Computing Center for the Arts and Design (ACCAD). One of the goals of the project was to explore dance as an object of investigation for cross-disciplinary research. The investigative process followed a fairly 'standard' flow from a dance as the study site/object, through data collection to analysis. The dance, "One Flat Thing, reproduced" (OFTr) is an ensemble dance, choreographed by William Forsythe, that examines and reconfigures classical choreographic principles such as different motions performed by dancers simultaneously. While the dance is choreographed and therefore somewhat 'controlled', the presence of significant improvisation makes many components unknown, even to the choreographer. Approaching this as a geographer, the parallel to time geography was striking, and the collected dataset gave us a unique combination of detailed space-time, activity, and communication information.

As novices to dance, it gave us a rare opportunity to be exposed to a spatio-temporal system that we had a fairly limited understanding of. I will give examples how visual annotations help even novice observers to start seeing shapes, patterns, and meaning in the dance. We now know, from our analysis, that certain themes link to other themes and that some are anchored in particular places or patterns of space and time. Still, we are far from what we would think of as a scientifically well described system. With that backdrop I reflect on how early stages of knowledge development is largely about observation, description, and some theoretical speculation. While this might be considered typical for humanities and social sciences we find the same pattern in much of the physical 'hard' sciences. For example, climate science have slowly built a consensus around what is affecting our climate and how. Early steps in this process was very much that of observation and description, rejection of initial hypotheses, and so on.

In a (1996) commentary Stan Openshaw speculated around research that on the one hand face a reality too hard to quantify but possible to describe, observe, formulate theory around, but hard analyze either because of limited data or fear of falling into simplistic positivism. On the other hand, was research that isolate subsets of reality down to analytically tractable components, but lacking the true complexity of interconnected systems and recursive feedback. In his commentary Openshaw argued for an analytical foundation that could better bridge the gap between simple but organized systems and complex but disorganized systems. For that middle-ground of 'organized complexity' he suggested fuzzy set theory as a possible scientific paradigm that could re-express qualitative theories in a tractable form. But the founding father of fuzzy set theory had even gone further to suggest that "...precise quantitative analyses of the behavior of humanistic systems are not likely to have much relevance to the real world societal, political, economic, and other types of problems which involve humans either as individuals or in groups." (Zadeh, 1973). If we concur, and in light of the contemporary understanding of human-environment systems as tightly integrated, it is questionable whether there is any prospect of developing truthful quantitative models of the environment. The suggested solution brought forward from this line of thinking is to acknowledge and provide support for the inherent imprecision in linguistic and other qualitative information. Qualitative knowledge representation has since emerged as a small but viable area for addressing the fusion of heterogeneous types of data in complex environments (Li, Dai, Dezert, and Smarandache, 2009).

Related efforts in the realm of formal category semantics, the proposition of a "Semantic Web", and the proliferation of ontology based approaches should be able to provide important building blocks in efforts to address qualitative computation. But most work in that area are unfortunately grounded in what I would (maybe incorrectly) characterize as mainstream computer logic. Even approaches to formal category semantics are mostly founded on crisp representations objects and relations. This is somewhat surprising given the fact that a major concern from the cognitive sciences has been that "...The gradation of properties in the world means that our smallish number of categories will never map perfectly onto all objects: The distinction between member and nonmembers will always be difficult to draw or will even be arbitrary in some cases [...] if the world consists of shadings and gradations and a rich mixture of different kinds of properties, then a limited number of concepts would almost have to be fuzzy." (Murphy, 2004). There are some notable exceptions though, and in my final example I will illustrate efforts that largely negotiate and bridge the gap between a discrete feature model of category semantics and a continuous geometric model.
This also provides another critical perspective on the use of the commonly taught system of measurement scales coined in the early 1940’s by Stevens (1946). The discourse around semantic similarity would support an argument that categories, or nominals, can be compared and evaluated in a graded and quantitative fashion. Nominals, and other linguistic terms, could then, in many situations, be regarded as an ordered and possibly even metric scale using appropriate semantic similarity metrics. It may be that semantic accuracy is possible to talk about in terms of semantic relations and associations. My final remark is then suggesting some potential implications of this for spatial analysis in general.

REFERENCES


