Advanced Engineering Informatics 38 (2018) 501–513

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

Game-based crowdsourcing to support collaborative customization of the definition of sustainability

Mazdak Nik Bakht[a], Tamer E. El-Diraby[b,*], Moein Hossaini[c]

[a] Dept. of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada
[b] Dept. of Civil and Mineral Engineering, University of Toronto, Toronto, ON, Canada
[c] Northwestern University, United States

ARTICLE INFO

Keywords:
Sustainability
Social media
Urban infrastructure
Computational linguistic
Machine learning
Crowdsourcing

ABSTRACT

Successful adoption and management of sustainable urban systems hinges on the community embracing these systems. Capturing citizens' ideas, views, and assessments of the built environment will be essential to this goal. In collaborative city planning, these are qualified and valued forms of partial knowledge that should be collectively used to shape the decision making process of urban planning. Among other tools, social media and online social network analytics can provide means to capture elements of such a distributed knowledge. While a structured definition of sustainability (normally dictated in a top-down fashion) may not sufficiently respond well to the pluralist nature of such knowledge acquisition; dealing with the unstructured community inputs, assessments and contributions on social media can be confusing. We can detect fully relevant topics/ideas in community discussions; but they typically suffer from lack of coherence.

In this paper, we advocate the use of a semi-structured approach for capturing, analyzing, and interpreting citizens' inputs. Public officials and professionals can develop the main elements (topical aspects) of sustainability, which can act as the skeleton of a taxonomy. It is however, the community inputs/ideas (in our case collected via social media and parsed), that can shape-up that skeleton and augment those topical aspects with adding the required semantic depth. In more specific terms, we collected tweets for four urban infrastructure mega-projects in North America. Then we used a game-with-a-purpose to crowdsource the identification of topics for a training set of tweets. This was then used to train machine learning algorithms to cluster the rest of collected tweets. We studied the semantic (finding the topics) of tweets as well as their sentiment (in terms of being opposing or supportive of a project). Our classification tested different decision trees with different topic hierarchies. We considered/extracted eight different linguistic features in studying contents of a tweet. Finally, we examined the accuracy of three algorithms in classifying tweets according to the sequence in the tree, and based on the extracted features. These are: K-nearest neighbors, Naïve Bayes classifiers and Support Vector Machines (SVM).

Respective to our data set, SVM outperformed other algorithms. Semantic analysis was insensitive to the depth/number of linguistic features considered. In contrast, sentiment analysis was enhanced when part of speech (PoS) was tracked. Interestingly, our work shows that considering the topic (semantic) of a tweet helped enhance the accuracy of sentiment analysis: including topical class as a feature in conducting sentiment analysis results in higher accuracies. This could be used as means to detect the evolution of community opinion: that topic-based social networks are evolving within the communities tweeting about urban projects. It could also be used to identify the topics of top priority to the community or the ones that have the widest spread of views. In our case, these were mainly the impacts of the design and engineering features on social issues.

1. Introduction

Decision making in green and smart cities is predicated on assuring sustainability on the one hand, and embracing more democratic project evaluation processes on the other hand. This gives rise to the formation of extended networks of decision makers/decision collaborators which involves the public communities actively and on an equal footing with official decision makers. The existence of informal sub-networks (of

* Corresponding author.
E-mail addresses: mazdak.nikbakht@concordia.ca (M. Nik Bakht), tamer@ecf.utoronto.ca (T.E. El-Diraby), moeinhosseini2020@u.northwestern.edu (M. Hossaini).

https://doi.org/10.1016/j.aei.2018.08.019
Received 21 March 2018; Received in revised form 14 June 2018; Accepted 31 August 2018
1474-0346/

sub-communities) in the process of network decision making is proven, in theory, to have a positive influence on reducing opinion clusters and reaching consensus among the official decision makers [39]. More specifically, an effective adoption of sustainable solution in the cities of today call for a more effective community engagement as such solutions normally require changes in behavior (in energy and water/resource usage patterns) and may impact some groups. Hence, community buy-in is essential to the success of these projects/policies. More interestingly, citizens possess valuable knowledge. Traditionally, this knowledge referred to "knowing the local context and conditions" or "knowing what they need and what works for them". On top of the local knowledge, today's "Net-savvy" generations possess significant abilities regarding technical issues too. In addition to a better education regarding sustainability, at their fingertips they have computer/smartphone applications, pervasive computing and cloud-based tools that allow them to examine data, develop scenarios, and collaborate in further evaluating them. In fact, in post-modernist planning theory, the role of the planner (decision maker) is no longer to develop solutions or push for a decision. Rather, it is to empower communities to define the problem and/or objectives; encourage social learning to enable reliable exchange of knowledge; and promote a healthy debate to support consensus or agreement about a community-created solution.

One of the essential challenges of green projects in the smart city is to develop 'user-inspired' and 'user-useful' approaches whereby citizens' knowledge is integrated with that of professionals (technical staff) [35,9]. At the core of harnessing citizens' knowledge is the contextualization of the very definition of sustainability. Sustainability has a context-sensitive definition. It is not easy to operationalize its high-level constructs into clearly-defined and (in some cases) measureable indicators. As far back as two decades ago, Dobson [12] surveyed over 300 attempts to define sustainability. Confusion and disagreement are typical and are an ever-present problem in any sustainability assessment exercise. With the expanding consideration of sustainability into all aspects of city management, these disagreements will only increase. For example, in studying urban sprawl (as one of the indicators/objectives of sustainable cities), Torrens [41] identified eleven characteristics of sprawl and used 42 different metrics. Lambert et al. [22] suggested fourteen criteria to evaluate civil infrastructure projects: create employment, reduce poverty, improve connectivity and accessibility, increase industrial/agricultural capacity, improve public services and utilities, reduce corruption/improve governance, increase private investment, improve education and health, improve emergency preparedness, improve refugee management, preserve religious and cultural heritage, improve media and information technology, increase women's participation and improve environmental and natural resource management.

One of the best solutions is to let the community define sustainability. This typically uses a mixed set of ethnographic methods including interviews and structured surveys. However, lately, analytics of community debates on social media have been suggested as a source for harnessing citizen knowledge—particularly about identifying the elements of sustainable development and in defining how to measure it. Classifier techniques are extensively used in this regards to analyze contents of community input to extract topics and then use these to create a model of the community-defined sustainability, with varying degrees of success/accuracy. While complete top-down definition of sustainability by public officials or professionals is rejected, a complete bottom-up definition will not be without problems. Heterogeneity in scoping and wording is the start [7]. Also, citizen input sometimes is tempered traditional practice, politics, and personal biases. Consequently, rigorous (and of course, sometimes, controversial) evaluation by professionals is still needed [10]. This is where the contribution of this paper is. In harnessing citizen views/models of sustainability, we adopt/propose the use of balanced approach between top-down and bottom-up definitions. Such a semi-structured approach is not new, many researchers have used it before. Professionals are to start with a basic model of sustainability (skeleton) with a set of major topics/issues. Then through interaction with communities, they refine such model—they fill in the contents of each topic/category. This approach had significant use with traditional communication modes (i.e. face-to-face interactions), but very limited use in the realm of social media communication modes. We hope that our approach will provide a working example for further development in this regards. The second contribution of this paper is to help enhance the clustering efficiency. Instead of professionals deciding what is the topic of a citizen contribution (a note on Facebook or a tweet), we used a game to get citizens to do that for a sample of tweets discussing four major urban projects. This is then used to train formal classifiers. This way, we complement input of professionals and citizens and use the crowd to explain what they mean when they share a comment or a note (what topic is being discuss).

The paper starts with a brief review of similar attempts for automation inside and outside the domain of public involvement for infrastructure projects. Then in Section 3 we explain our methodology for data collection, selection of features and algorithms used to train the classifiers, and the measures used to evaluate their performance. Section 4 provides details on data collection and preparation. Section 5 discusses the use and analysis of machine learning algorithms and linguistic features in clustering tweets based on their semantics and sentiment. It also discusses the relevance and accuracy of the used algorithms. Finally, concluding remarks are presented in Section 7.

2. Related works

Stakeholder management in infrastructure planning and construction is principally interested in detecting subnetworks of citizens, relying on, or impacted by an infrastructure project; their vested-interests in the project; and their positions (in terms of being in-favor or against) with respect to the decisions made in that project [32]). It is globally accepted that the public opinion is shaping-up the attributes and definition of products and services they use [8]. The impact of sub-networks' discussions on the opinion dynamics of decision makers is studied in different cases, such as military Command and Control systems with single and multi-level threshold opinion dynamic models [40]; and hierarchical formal organizations of different constructs [39]). By studying different types of bottom-up informal networks, Song et al. identified three main attributes of the informal sub-networks (namely: agents' tolerances, scale and number of links) to have the highest level of impact on the officials' opinion. Distilling context-related knowledge (from the citizens) can help decision makers set matching decision goals and evaluate performance of proposed solutions. This requires collecting and processing views from the citizens who interact with the physical infrastructure, and use the service provided by it, on a regular basis. The use of citizens as knowledge sensors includes two main streams: hiring citizens to collect data (or to solve complex problems) and using citizens to validate the knowledge extracted through engaging other citizens. The former includes methods such as citizen sensor networks (hiring the citizens or using volunteers for collecting data). This includes, for example, crowdsourcing for inspection and assessment of facility attributes (type, age, dwelling, etc.) or collection of infrastructure defects/condition data [38]; hydro-geologic evaluations and monitoring stream stage [25]; sensing earthquake [3]; etc. Extensions of such applications exploit citizen science: the ability of lay people to find innovative solutions to technical problems. Some of the new available applications have resulted in solving complex problems through human-based computation for tasks such as computational fluid dynamics simulation [43]; photo tagging of building and infrastructure damages [44]; map refinement [20].

The second stream, on the other hand, aims to categorize or detect patterns in aggregated content generated by ordinary citizens or the footprint of their behaviour in online social networks and social media. Data collected from citizen sensor networks can be translated into