Empowering the Human: A Crowd-sourcing Perspective

Tarek Abdelzaher, Department of Computer Science

Extended Abstract: Cell-phones have one of the highest penetration ratios in developing countries, compared to other technological artifacts such as computers, Internet, air conditioning, and personal vehicles. This penetration is on a rapidly increasing slope. For example, India, one of the world's fastest growing Wireless markets, is estimated to have 729.57 Million mobile phone subscribers as of November, 2010. This is significantly more than the total number of households. It is projected to have 1.159 billion mobile subscribers by 2013. The above statistics suggest that mobile phones can be a powerful tool for reaching the bulk of the population both for understanding needs and disseminating solutions.

A key impediment to efficient post-harvest crop distribution in developing countries lies in the lack of streamlined supply chains, due in part to the lack of technology that enables efficient monitoring, transportation, storage, and coordination along the different stages. Cell-phones are an exception in terms of wide-spread availability, yet their potential in counteracting the above monitoring and communication deficiencies remains largely untapped. The above motivates developing a "virtual information highway" for post-harvest supply chains, based on widely available cell-phones. It will leverage the most extensive natural resource available on the supply chain: the human in the loop. The virtual information highway will rely on crowd-sourcing, which refers to the involvement of the general population in data collection and feedback. In this case, willing participation of self-interested parties (including farmers, distributors, warehouse workers, and other supply chain actors) can enable robust monitoring and management of post-harvest supply chains. The proposed service, by collecting information and offering feedback at a large scale, will significantly improve awareness of bottlenecks, inefficiencies, needs, and causes of loss. It will also improve the ability to coordinate globally to streamline the process using only the resources available. Finally, it may shed light on desirable incentive structures that can prod actors along the supply chains to consider various behavioral patterns that improve cooperation and minimize global loss.

Abdelzaher and collaborators at the Department of Computer Science bring expertise on human-centric sensing (crowd-sourcing), sensor and communication networks, information processing, and social networks, all of which can be brought to bear on the above problem. As a pilot project, Abdelzaher proposes to build an automated information collection and management service that allows a provider to prompt large groups of individuals to share data via their cell-phones on selected issues of concern (in the form of text, voice, or images, depending on cell-phone capabilities). The service back-end then performs data cleaning and analytics to demonstrate the viability of extracting reliable information from the large number of individually potentially-unreliable or biased sources that might join the data collection campaign. While the ultimate objective of such a system would be to collect actionable information from farmers and supply chain workers on bottlenecks and inefficiencies that cause post-harvest loss, as a test of this system it will be used locally on campus for a few chosen data collection campaigns. For example, students may be asked to share (in voice, text, or images) instances of observed energy consumption on campus that they believe is a candidate for saving. The system will then extract a ranked list of recommendations for a greener campus. Other campaigns may involve reporting traffic congestion or maintenance needs in public areas. The project will serve as a research vehicle and proof of concept to investigate and demonstrate the following:

- *Viability of extracting accurate information using crowd-sourcing:* Crowd-sourcing relies on people, as opposed to well-calibrated sensors. Human perception is subjective and biased. The project will investigate the degree to which crowd-sourcing data can be cleaned to yield useful information. Fact-finding algorithms (e.g., Google's Page Rank that estimates credibility of web sources) will be experimented with to derive credible information.
- *Scalability:* To illustrate how the above algorithms scale, we shall use large volumes of Twitter data (note that all tweets are publicly available) to reliably reconstruct sequences of reported events, extracted from large human populations; for example, the catastrophic events of the Japan Tsunami. Extracted events will be compared to ground truth obtained from international media. The exercise will demonstrate scalability of accurate fact-finding.
- *Viability of data fusion from voice, text, and images:* Heterogeneous information in the form of text, images, and voice may be collected and will need to be consolidated. Abdelzaher currently leads an ARL-funded project on heterogeneous data fusion from unstructured sources in military applications. Algorithms developed in this project will be re-tooled and tested to demonstrate fusion of heterogeneous crowd-sourcing data sources.
- *Incentives:* Incentives and recruitment issues will be addressed for adequate representation of target populations.

The aforementioned proof-of-concept effort will help establish the viability of developing a virtual information highway based on crowd-sourcing. If successful, this effort will pave the way for a larger engagement that applies the proposed mechanisms to the problem of understanding and reducing post-harvest loss.