Spying with the Crowd

Malay Bhattacharyya
Department of Information Technology
Indian Institute of Engineering Science and Technology, Shibpur
Howrah – 711103, India
E-mail: malaybhattacharyya@it.iiests.ac.in

Abstract

The cooperative power of crowd workers has been recently shown to be effective in solving large-scale diverse tasks. We elaborately study one such application of collaborative crowdsourcing. In this paper, we discuss the viability and different facets of crowdspying, activity of spying with the power of crowd workers. This type of citizen involvement has diverse applications that also includes large-scale social security. We provide some highlights into its different varieties and underline how to build successful models.

Introduction

Recent research in human computation has witnessed a rapid popularity of crowdsourcing models. It has been established that a large pool of cooperative crowd workers might be superior than limited experts in performing large-scale diverse tasks. Such models have received substantial attention in the recent years for their distributed yet united power of problem solving (Kittur et al. 2013). Based on the working behavior, crowdsourcing can take either an explicit or implicit route. Explicit crowdsourcing allows users to work together to evaluate, share and build different specific tasks, while in implicit crowdsourcing the users solve a problem as an effect of something else that was done by them. Here, we propose a novel application of spying with the crowd, which combines explicit and implicit crowdsourcing. Spying with the crowd, or crowdspying, is to take the help of the crowd workers for exploring the truth or undercover operation. It is principally different than spying on the crowd (Fielding and Cobain 2011), rather it spies with the help of the crowd. DARPA's Red Balloon Challenge was the first such attempt and the models of the winning team (Pickard et al. 2011) and the "I Spy a Red Balloon" team (Tang et al. 2011), who stood second, shows the power of crowdspying in a limited setting. We highlight some models that works on social network platforms for crowdspying purposes. We discuss the varieties of crowdspying and also study the viability of such models. We highlight some open problems and discuss about some solutions to address these issues.

Preliminaries

Crowdspying is a kind of collaborative crowdsourcing activity. In crowdspying, both the explicit and implicit nature of actions are combined. We assume that crowdspying is confined to an online social network. Then, we can formalize any arbitrary crowdspying activity with the quadruplet $(\mathcal{N}, s, \mathcal{P}, \mathcal{R})$, where $\mathcal{N}$ is the node set, $s \in \mathcal{N}$ is the node to spy on (spied node), $\mathcal{P} \subseteq \mathcal{N} \times \mathcal{N}$ is the set of arcs and $\mathcal{R} \subseteq \mathcal{N}^{|\mathcal{N}|}$ is a private relation between the nodes. A node is basically a participating entity in the crowdspying operation (therefore it also includes the spied node). The private relation R is used to denote the connection between multiple nodes (like hyperarcs). For example, in some social networking sites friend list denotes a private (closed) connection between the profiles through which information are shared. Therefore, the most significant nodes are the neighbors of the spied node. Let us define the following.

Definition 1 *A node having an arc with the spied node is called the* leak node*.*

The *leak node*s might be of two types – having directed or undirected arcs with the spied node. Our goal is to set up a path of spying that constitutes of multiple arcs and nodes but not the spied node (strict constraint) or the leak nodes (flexible constraint).

Categories of Crowdspying

We have already seen that there can be two versions of crowdspying (with strict or flexible constraints). Other than this, there can be a number of variants of the crowdspying activity. Broadly, it can be categorized into various types based on different factors. These factors and how they categorize crowdspying are described below.

Based on information sharing

Based on how the information is shared between the nodes, crowdspying can be of two different types.

Definition 2 Closed crowdspying *is an activity in which relationships exist between the nodes participating in crowdspying, i.e., the private relation set R is non-null.*

E.g., the crowdspying activities that might happen through the social networking platforms like Facebook or Google+ are *closed crowdspying*.

Definition 3 Open crowdspying *is an activity where the nodes participating in crowdspying are independent of any relation, i.e., the private relation set R is null.*

E.g., *open crowdspying* might happen via social networking platforms where information are publicly accessible.

Based on control

The crowdspying can be operated by an administrator and this invokes the following further subtypes.

Definition 4 *The crowdspying activity in which the information flow is guided by a mechanism design adopted by an administrator is a* controlled crowdspying.

Definition 5 *The crowdspying activity which is not guided by any mechanism design is an* automatic crowdspying.

Any arbitrary crowdspying activity is really hard to manage in an organized way towards a specific goal.

Applications of Crowdspying

The major advantage of crowdspying is its zero cost as it involves crowd volunteers. We describe the potential directions of crowdspying applications. In each case, we point out how crowdspying is distinct from the conventional spying. Spying is not a social activity, rather it emerges from two basic necessities. Primarily, it is nothing but a professional activity where payment is given for spying – may be by a person or from the Government. Otherwise, it is used to fulfill the personal demands like collecting information. Crowdspying is mainly demanding for its social cause. Several such applications are highlighted below.

- **Filtering recruitment applications:** Any professional (or even personal) information can be verified through crowdspying. It can be used in the recruitment process or even for the admission to graduate programs.
- **Assisting in social security:** This might include identifying the intruders, terrorist activity or even search for victims. It is easily understandable that the crowdspying for social security should be closed, and controlled by the Government based on robust mechanism designs.
- **Tracking human trafficking:** Spying to recognize human trafficking is highly demanding for its distributed power of information sharing. This type of crowdspying is useful for rapid transfer of information and therefore better to be open.

Threats in Crowdspying

The main concern in developing crowdspying models is the threats involved in it. As it deals with the criminals or bad people (who suppress information) and might include the leak nodes, the threats of getting counterattacked becomes nigh. The depth-breadth expansion of the information flow through a network of crowdspying activity makes it accessible (even for the closed crowdspying model) to many people. So, the inclusion of clever mechanisms to spy on a node that is also a part of the network is a major area of further research. Note that, the crowd is public and spying is a private activity. One solution to this problem might be keeping the spied node out of the participating nodes. However, the problem naturally occurs even if we discard the spied node because leak nodes are always there.

Figure 1: The WeSpy interface for studying crowdspying.

Crowdspying Implementation

Practical implementation of crowdspying is a really a challenging task. Other than threats that we already discussed, another major problem is that how to make the system run. A solution could be using the social networking sites. The social networks can propagate information for some implicit purpose that might lead to crowdspying in the background. To test the feasibility, we designed an interface, named WeSpy (see Fig. 1), for the analysis of a basic open crowdspying model involving 101 people. But the major problem appears to create interest in the crowd to volunteer for crowdspying.

Challenges and Concluding Remarks

Our primary results show some interesting patterns of activity. The crowd workers appear to be more interested in public level crowdspying but not for jobs involving higher threats. Again, we found more number of males are interested in these activity. Studying the overall crowdspying network might provide additional information about the nodes that are leaders in this activity. The pattern of *open crowdspying* and *closed crowdspying* might also provide important information. But the major challenge remains to be attracting crowd volunteers and managing threats. The limitation of this study is a small-scale analysis but it provides some important insights. Overall, our approach showed that crowd can be used effectively for large-scale spying activity but with careful mechanism design.

References

Fielding, N., and Cobain, I. 2011. Revealed: US spy operation that manipulates social media. London: theguardian.

Kittur, A.; Nickerson, J. V.; Bernstein, M. S.; Gerber, E. M.; Shaw, A.; Zimmerman, J.; Lease, M.; and Horton, J. J. 2013. The Future of Crowd Work. In *Proceedings of the CSCW*, 1301–1318. ACM Press.

Pickard, G.; Pan, W.; Rahwan, I.; Cebrian, M.; Crane, R.; Madan, A.; and Pentland, A. 2011. Time-Critical Social Mobilization. *Science* 334(6055):509–512.

Tang, J. C.; Cebrian, M.; Giacobe, N. A.; Kim, H.-W.; Kim, T.; and Wickert, D. B. 2011. Reflecting on the DARPA Red Balloon Challenge. *Communications of the ACM* 54(4):78–85.