Crowdtrust
Group Project

Giovanni Charles Adam Fiksen Ryan Jackson John Walker

January 4, 2013

1 Introduction

1.1 An Informal Discussion of Crowdsourcing and the problems it faces

Our project concerns *crowdsourcing* 'the book' would describe crowdsourcing as the principle of obtaining an accurate and appropriate result by having many different contributors performing a task but we'd like to start with a story which we believe encapsulates the idea of crowdsourcing in its most positive and useful light.

In January 2009 Timothy Gowers (Fields Medal winner and avid blogger) used his blog to post a striking question *Is massively collaborative mathematics possible?*. He posted a difficult and unsolved mathematical problem he was particularly interested in and invited people to contribute to its solution in the comments section. The project initally got off to a slow start but once the ice was broken the comments flooded in. 37 days , 27 contibuters and 800 comments later Timothy Gowers was able to announce that not only had they solved the original problem but they had also solved a harder generalisation of it, he called his experiment the *Polymath Project* .

This is a nice example of how crowd sourcing can be used to combine the skills of many individuals and produce answers to a complex problems, but there are many motivations to outsource your task to the crowd:

- *Computational Difficulty:* Timothy Gowers provided a nice example of a computationally difficult problem It is extememly unlikely you would be able to write a Java program or use Wolphram Alpha to produce a complex mathematical proof, some problems require what we like to call the 'human touch'. Problems which require the human touch are in no way confined to the realms of complex mathematics identifing an unknown bird in a picture for example would prove quite difficult on a computer you may need access some a program like Google Goggles but even then you would probablly need a good quality photograph, whereas one avid bird enthusiast in the crowd might be able to easilly identify the bird and reuturn the correct answer.
- *Saving Time:* In 2009 aviator Steve Fosset crash landed on an island off Antigua, his friends back in America knew they had a very small and time critical window to find him alive and they had little faith in the current search and resuce operations. They organised for satellite image to be taken of the island and the images were passed to a crowd who were asked to

identify foreign object which could be potential crash sites. This is a nice example of how the crowd can be used to literally cover a large ammount of ground in a small ammount of time. Not all examples of saving time are quite this dramatic though, image tagging is an extemely arduous task for an individual or small group of people to perform and tagging a relatively large set of images could take weeks or even months, outsourcing this to the crowd could have the job done in a number of hours.

- *Saving Money:* Time is money as they say and this goes hand it hand with the point above. If you have to pay a team of high salary computing professionals to tag images for your project when you could be paying them to write code this is not cost effective, however passing this job off to a crowd of lower paid people could potentially save you a lot of money.

- *Reaching A Willing Audiance:* Unless it's something they enjoy the fact is that people are not willing to work for free, this is why getting the general public to do things such as complete surverys can be difficult as a large number of people will simply not want to do it. The crowd memebers will be inventivised to perform work and as such will be more likely to complete your survey.

This is all well and good in theory but it leaves us with a number of problems, an introduction to these problems is provided below but they are discussed and addressed in much greater details at various stages throughout the report:

1. *How do we get these problems to the crowd?*
   It is unlikely many people with a problem to be solved would want to go to the trouble of creating a crowd themselves as this would be comparitive to the complexity of the problem itself therefore there is a need for a third party crowd management system.

2. *What problems can we ask the crowd?*
   Specialised crowds have been successful and certainly have their uses for instance www.stackoverflow.com can be thought of as a crowd specialising in the solution of computing problems, however it is unlikely I would be able to find a specialised crowd to indentify bird pictures or to search satellite images for crash sites, therefore I need access to a generalised crowd able to adapt to and solve a wide variety of problems. The crowd management party therefore needs to provide the ability to ask a wide variety of questions and the ability to easily encorporate new questions in response to new technology.

3. *How many people do we ask, who do we ask and how can we trust what they say?*
   Crowd members will be a representitive sample of the general population, some will be brighter than others and some will be willing to put in more effort than others based on this you can place all annotators on a scale of trustworthiness which rates how much you believe an answer they give you. This rasises the question of 'how many people do I ask?', is consulting a small number of very trustworthy people better than a large number of non trustworthy people?, However I can't simply ask the same subset of people over and over again as workload will build up and my answers will be delayed. If I have a 'specialist' question do I direct it to someone with knowledge of that specialism? Clearly a sophisticated algorithm is needed on the crowd management side to address these problems.

A solution to these problems provides us with the basis for our project we are to:

Design and implement a framework that allows the optimum number of contributors to be selected on the basis of their trustworthiness for a desired accuracy of the outcome/result and that evaluates the trustworthiness of contributors on the basis of the accuracy of the results they provide.

1.2 A Formal Specification of Our Objectives

1.3 A Formal Discussion of Our Achievements

2 Design and Implementation

3 Evaluation

4 Conclusion and Future Extensions

5 Project Management

6 Appendix

References