On the Dynamics of Micro- and Macro-task Human Computation Markets

Lesandro Ponciano, Francisco Brasileiro
Department of Computing and Systems
Universidade Federal de Campina Grande

July, 2013

1 Introduction

The last few years have witnessed a significant growth in human computation systems. One kind of system that has received more attention from human computation users is the so-called *online labor market* platform. Such platforms are human computation systems that act as an interface between *requesters*, users who submit human computation tasks to be executed, and *workers*, users who perform human computation tasks as paid work [3, 1].

Regarding task granularity, online labor market can be broadly divided into two classes, namely: micro-task platforms and macro-task platforms. Examples of micro-task platforms currently in production are: mturk.com, microworkers.com, microtask.com, and minifreelance.com. Example of macro-task platforms currently in production are: odesk.com, elance.com, and guru.com. These distinct task granularities also lead to a distinct profile of workers in the two classes of platforms. Usually, macro-task platforms gather workers expert in some kind of skills, while micro-task markets gather workers able to perform tasks that require only basic human abilities.

Motivation: With the growth of human computation field, new platforms are emerging and ever more requesters are using them to execute their human computation tasks [4, 2, 1]. However, little is known about the behaviors of crowd of workers and requesters in the platform and the typical workload they generate. Furthermore, little is known about how such workload differs between micro- and macro-tasks human computation platforms.

Goals and Contribution: This study presents an in-depth analysis of human computation workload in online labor markets. It is based on an extensive dataset collected from Mturk.com, which is a micro-task online labor market, and from Elance.com, which is a macro-task online labor market. This dataset consists of more than 2.6 millions of jobs resulted in almost one TeraByte of data that cover a 4-month time period. By using this dataset, we provide a statistical analysis of online labor market workload, which focuses on requesters, workers, and tasks characteristics in both micro- and macro-task human computation markets.

Paper Outline: The remainder of this paper is organized as follows: In section 3, we give a high-level description of the data set used in our online labor market characterization. In section 4 we present our online labor market characterization. This is done from three different perspectives. The first characterizes how individual requesters interact with the system. This is a requester's perspective. The second focus on crowd worker characteristics. This is a crowd's perspective. The third characterizes the aggregate of tasks patterns of all requesters. This is a platform view of the workload. In section 5 we present our discussion and implications for design. Finally, in section 6, we present the conclusions of this work.

2 Background and Related Work

2.1 Background

Before presenting the human computation components, it is important to establish a distinction between jobs and tasks. A job is set of one or more tasks (or Human Intelligence Tasks, HITs). In micro-tasks markets, a job is constituted of several small tasks and in macro-tasks market a job is constituted of only one task. The jobs submitted by all requesters are made available in the platform in *job boards*. In such boards, workers can choose the task they want to perform. Usually, in micro-task market one task is assigned to a worker as soon as he/she chooses it, while in macro-task market several workers may choose the same task and the requester needs to choose the worker to which the task will be assigned. Figure 1 shows an example of a job constituted of micro-tasks (Fig. 1(b)) and a job constituted of one macro-task (Fig. 1(a)).

(a) Example of a job in Elance macro-task market

(b) Example of a job in Mturk micro-task market

Figure 1: Examples of jobs description in Human Computation Market

We analyze online labor markets workload from three different perspectives: requesters, crowd of workers, and platform. These three perspective are shown in Figure 2. Platforms

are human computation systems that receive jobs submitted by requesters and make such jobs available to the crowd of workers. The aggregated of jobs submitted by requesters generate the demand for skills in the platform. Crowd workers is the aggregated of human workers in a platform; it constitutes the supply of skills in the platform. Requesters act in the platform submitting jobs to be executed.

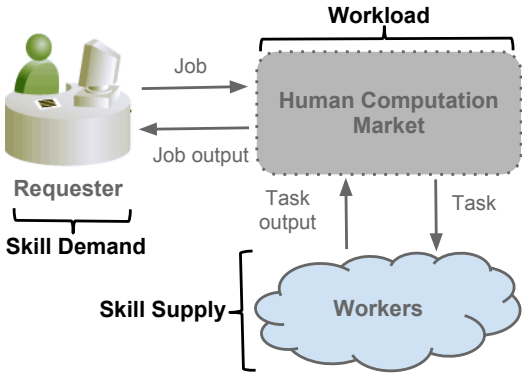


Figure 2: Type of work demand

2.2 Related Work

Only a few studies have focused on analyzing online labor market performance [1]. Ipeirotis [1] presents a general analysis of the micro-task market Amazon Mechanical Turk (Mturk) data. In this study we analyze Mturk workload focusing on requesters, crowd of workers and platform workfload. In addition, our study allows us to establish a comparison between micro and macro-task markets.

3 Materials and Methods

3.1 Data Collection and Overview

We collected data from two online labor market platforms: Mturk (mturk.com), and Elance (elance.com). Mturk is a micro-task human computation platform and Elance is a macro-task human computation platform. We focused on collecting data which reflect the perspective of requesters and the crowd of workers in the public web page of the platforms. In both platforms, the data were collected at every 2 minutes. The collected data differ in the platforms as following:

- *Mturk:* We collected the following job data: job id, requester id, reward, number of tasks (HITs), and required qualifications. We can compute the submission time, and the evolution of the number of HITs available for execution over time. In Mturk, there is no information available about workers individually. Thus, we collected general information of the qualifications of the crowd of workers. We collected the following data about crowd of workers qualification: qualification id, title of the qualification, number of qualified workers.

- *Elance:* We collected the following tasks data: job id, requester id, reward, number of proposals, and skills required. By collecting such data at every 2 minutes, we can compute job submission time, and the evolution of proposals over time. Regarding crowd of workers, we collected the following data: worker id, worker name, worker reputation, the amount of earnings, workers skills, and geographical location.

Table 1 shows a summary of the data. This table shows that the platforms are significantly different in terms of the number of active requesters and the number of jobs submitted. Compared to Elance, Mturk shows more submitted jobs, but less active requesters.

Table 1: Summary of the datasets

Characteristic	**Mturk**	**Elance**
Start date	2012-10-24	2012-11-01
Duration (days)	120	120
#Unique requesters	3,523	54,946
#Jobs submitted	2,372,255	253,209

3.2 Characteristics

Activity level of requesters: Requesters act in the systems submitting jobs to be executed and receiving their results. In this study, we focus on characterizing the activity level in which requesters interact with the system by using two metrics: *requesters activity duration* and *resources demand*. Activity duration is the number of days that one requester has submitted at least one job. Requester's demand is the number of tasks that he/she has submitted. These metrics allow us to analyze the demand for skills in the platforms.

Skills of crowd workers: Workers act in the system executing task. Crowd of workers is the aggregated of workers in a platform. We focus on analyzing the distribution of skills in the crowd of workers. This metric allows us to analyze the skill supply in the platforms.

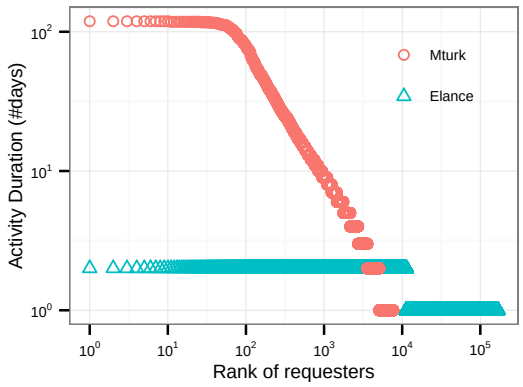
Platform workload: Platform workload is the aggregated of all jobs submitted by all requesters. We focus on two characteristics of the workload: *required skill* and *time cycles*. Time cycle of activity indicates how likely it is that requesters submit more tasks at certain hours of day or day of the week. A standard to identify daily cycles is to measure the proportion of tasks that were submitted at each hour of the day (*daily cycle*) or at each day of the week (*weekly cycle*).

4 Analyzing Requesters, Crowd of Workers and Platform Workload

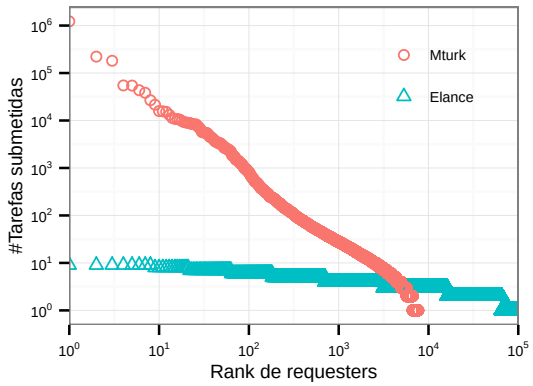
4.1 Requesters

Requesters Activity: Figure 3 shows the distribution of requesters activity in terms of the number of days that they submitted at least one task (Fig. 3(a)) and in terms of the amount of jobs they submitted (Fig. 3(b)). Each point in the image corresponds to one requester.

In both figures, the axes are in log-log scale. The horizontal axis shows requesters' rank from the highest to the lowest level of activity. In Figure 3(a) the vertical axis shows the activity duration in number of days, while in Figure 3(b) the vertical axis shows the number of jobs submitted.



(a) Number of days



(b) Number of submitted Tasks

Figure 3: Distribution of requesters' activity

Regarding the comparison between the platforms, our results indicate that (i) in the Elance platform requesters show lower activity than those in Mturk, and (ii) Elance shows more active requesters than in Mturk. In Mturk, 44 (1.25%) requesters submitted at least one task in every 120 days of our monitoring. Both the activity duration and number of submitted jobs by requesters in Mturk are skewed. In general, Mturk requesters submit more jobs and more often than Elance requesters.

Skill Demand: Figure 4 shows the distribution of the demand per skills in Mturk and Elance. In this figure, skill demand is measured by number of jobs. The axes in this figure are in log-log scale. The vertical axis shows the demand for skill. The horizontal axis shows

skills' rank from the highest to the lowest demand. Each point in the image corresponds to one skill. This figure shows that in both platforms, although a number of skills are demanded, the larger proportion of the demand concentrates in few skills.

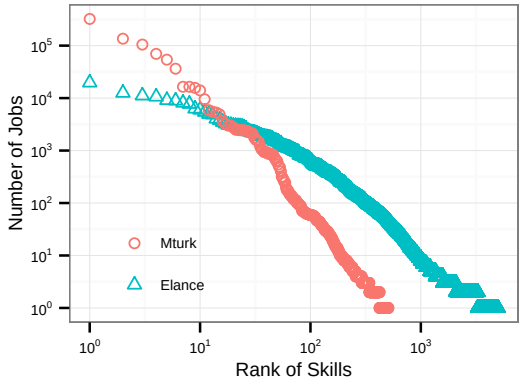


Figure 4: Skills distribution in the workload

4.2 Crowd of Workers

Skill Supply: Figure 5 shows the distribution of skill supply in the platforms. In this figure, skill supply is measured by the number of qualified workers. The axes in this figure are in log-log scale. The vertical axis shows the skill supply. The horizontal axis shows skills' rank from the highest to the lowest supply. Each point in the image corresponds to one skill.

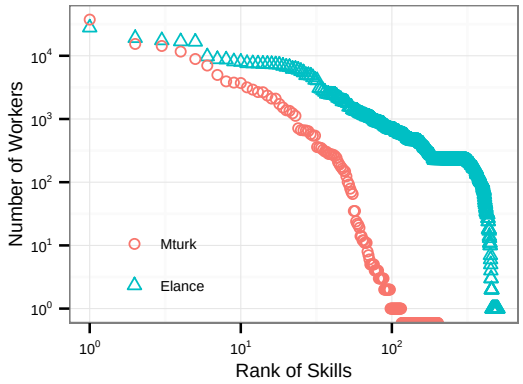


Figure 5: Skills distribution in the Crowd of Workers

As shown in this figure, in both platforms the distribution of worker per skills is not very concentrated. This indicates that the crowd of workers gathered by the platforms is specialized in a specific skill. Furthermore, there is a diversity of qualified workers in each skill.

4.3 Platform Workload

Job arrival patterns: Figure 6 is a boxplot of the number of new jobs that arrive in the platform at every 2 minutes. This figure shows that job arrivals in Mturk is higher than in Elance, although the number of active requesters in Mturk is smaller than in Elance. At every 2 minutes, an average of 30 new jobs arrive in Mturk, while an average of 14 new jobs arrive in Elance.

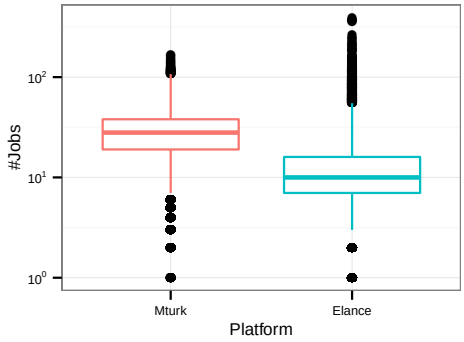


Figure 6: Number of jobs that arrive at each 2 minutes

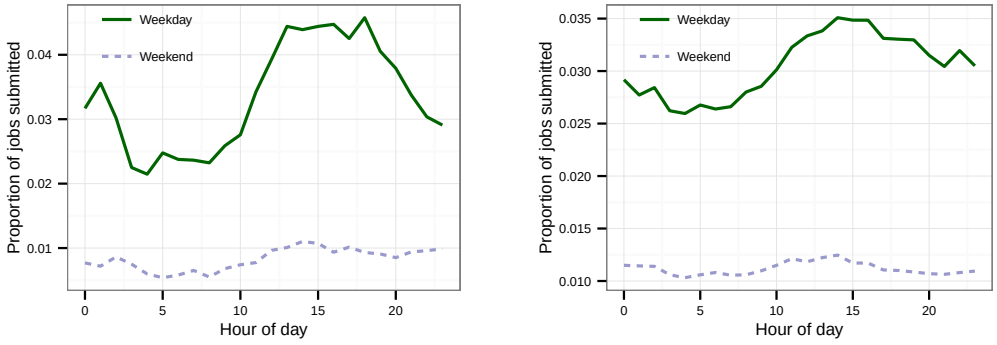
In micro task-markets, the time that workers spend searching for tasks in the job board may be equivalent to or even larger than the time that some tasks require to be executed. Thus, workers tend to spend little time seeking task information such as required skills and reward. This behavior associated with the higher dynamics of job boards may cause the following behaviors: (i) either workers do not choose properly the job or they spend more time searching for a suitable job; and (ii) jobs tend to experiment starvation as they go to the end of the job board.

Daily Cycles: Daily Cycle of Activity indicates how likely it is that requesters submit more tasks at certain hours of the day. A standard to identify daily cycles is to measure the proportion of tasks that were submitted at each hour of the day. Figure 7 shows this analysis. The horizontal axis shows the hours of the day. The vertical axis shows the proportion of all tasks. The lines distinguish weekdays from weekends.

It is clear that there is a pronounced effect of the daily cycle on both platforms. While variations occur, both platforms exhibit more job submission between 10:00 and 20:00 hours and peak near 15:00 hours.

Weekly Cycles: Figure 8 shows the weekly-cycle analysis. Weekly Cycle of Activity indicates how likely it is that requesters submit more tasks at certain days of the week. A standard to identify daily cycles is to measure the proportion of tasks that were submitted at each day of the week. Figure 7 shows this analysis. The horizontal axis shows the days of the week. The vertical axis shows the proportion of all tasks. The lines distinguish the platforms.

It is clear that there is a pronounced effect of the weekly cycle on both platforms. While variations occur, both platforms exhibit more job submissions in weekdays than weekends.



(a) Elance

(b) MturkTrace

Figure 7: Daily Cycle

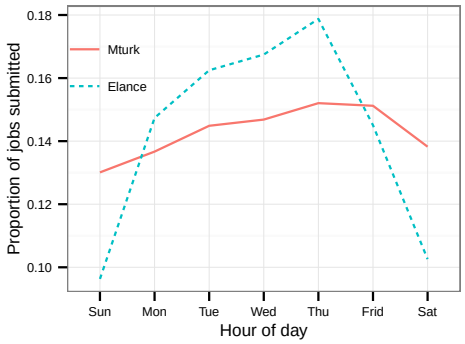


Figure 8: Weekly Cycle

Both platforms exhibit peak of job submissions on Fridays.

5 Discussion and Future Work

In this study, we found that online labor markets are heterogeneous mainly in terms of workers' skills and requesters' activity levels. These markets also differ in the proportion of skilled workers to perform each type of task and in the dynamics with which requesters submit their tasks. These market characteristics generate variations in the amount of attention a job receives when submitted to be executed in a given market. Thus, a challenge in the current human computation markets is to provide mechanisms to estimate the quality of service that can be expected in each market at the time that the job is submitted. This information can take the requester to make inappropriate choices in terms of job design and platform selection, aiming at reducing the waiting time to get workers enrolled to perform the tasks, in micro-task platforms, or getting a higher diversity of workers willing to perform the task, in macro-task platforms. Note that, studies in this context need to distinguish

between the impact of market factors and the impact of job design factors on job attention.

Market factors: As market factors, we consider the skills of workers and the ratio between supply and demand for these skills. It seems that market factors can impact job attention. For example, a system may consider the current status of the market in terms of skill demand and supply and the typical daily and weekly cycles in order to estimate the public of workers the job will attract.

Job design factors: Once market factors are isolated, it is important to understand how to define the job parameters, such as: financial compensation for the execution of each task, the number of tasks that composes each job, and the amount of time required for the execution of each task. A system can suggest to the requester values for these parameters in order to increase the attractiveness of tasks. These values can be suggested based on historical information about jobs submitted to the platform and the current state of supply and demand for jobs in the system.

6 Conclusions

In the present study we conducted an in-depth analysis of human computation workload in micro- and macro-task human computation markets. Our study is based on an extensive dataset collected from Mturk.com, which is a micro-task market, and from Elance.com, which is a macro-task market. Our study shows that the behavior or workers and requesters is different in both classes of online labor markets platforms. In micro-task market, requesters seem to be more active in terms of both the number of jobs submitted and the number of days in which at least one job is submitted. The platforms also show daily and weekly cycles in task submission. When discussing our results, we put into perspective opportunities for future work that takes into account our results to improve task assignment, task design and platform selection.

References

[1] Panagiotis G. Ipeirotis. Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students*, 17(2):16–21, December 2010.

[2] Alexander J. Quinn and Benjamin B. Bederson. Human computation: a survey and taxonomy of a growing field. In *Proc. CHI '11*, pages 1403–1412, Vancourver/BC/Canada, 2011. ACM.

[3] Man-Ching Yuen, I. King, and Kwong-Sak Leung. A survey of crowdsourcing systems. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, BPM'11, pages 766 –773, oct. 2011.

[4] Haoqi Zhang, Edith Law, Rob Miller, Krzysztof Gajos, David Parkes, and Eric Horvitz. Human computation tasks with global constraints. In *Proc. CHI '12*, pages 217–226, New York, NY, USA, 2012. ACM.