

Secondo Progetto Social Computing

Prima parte

A.A. 2021/22

M. Giunta¹

S. Bibalo²

G. De Nardi ³

Gennaio 2022

¹Marco Giunta 147852 giunta.marco@spes.uniud.it

²Simone Bibalo 143184 bibalo.simone@spes.uniud.it

³Gianluca Giuseppe Maria De Nardi 142733@spes.uniud.it

1 Introduzione

In questa prima parte del secondo progetto, abbiamo realizzato un task di crowdsourcing usando la piattaforma Amazon Mechanical Turk (da qui in avanti indicata come Mturk). In realtà, stiamo usando la versione *sandbox* di Mturk, piattaforma di sviluppo utile per testare i propri task ma dove i worker non vengono pagati realmente.

Per generare il task abbiamo utilizzato il software Crowd_Frame. Questo software semplifica la creazione dei task, sfruttando la piattaforma Mturk solo per la ricerca dei worker e per il loro pagamento. Infatti, il codice vero e proprio del task è stato creato all'interno di un bucket Amazon S3 e su Mturk viene inserito solo il link per raggiungerlo. Questo permette maggior flessibilità nella creazione dei task, visti i limiti dell'interfaccia di Mturk.

2 Crowd_Frame

2.1 Configurazione

Come impostazioni generali di Crowd_Frame abbiamo deciso che ogni worker deve impiegare almeno 2 secondi per rispondere alle domande del questionario o per analizzare ogni singolo tweet. Inoltre abbiamo dato la possibilità al worker di ripetere il task fino a 10 volte, nel caso non superi il controllo qualità.

2.2 Tokens

Sono state generate 15 coppie casuali di stringhe da usare come *input* e *output tokens* per gli HITs.

2.3 Questionario

Per raccogliere informazioni sull'uso che il worker fa di Twitter, abbiamo aggiunto un questionario con quattro domande a risposta multipla:

1. Per cosa utilizzi Twitter?
2. Quanto tempo passi in media su Twitter?
3. Di media ogni quanto hai una interazione attiva (twitt/retwitt)?
4. Hai mai comprato qualcosa dopo averla vista su Twitter?

2.4 Dimensioni

Il task è formato da 15 HITs: in ognuno, il worker deve analizzare 5 tweet. Per quanto riguarda la dimensione a piacere, abbiamo scelto una domanda a risposta multipla: al worker viene chiesto se "l'argomento del Tweet riguarda il campo accademico".

2.5 Tweets

Sono stati selezionati 15 tweet in lingua inglese dalla lista di quelli usati nel primo progetto di Social Computing. La scelta non è stata casuale: abbiamo evitato di selezionare *Reweeet* o *Quote Tweet*, in modo da avere solo tweet scritti direttamente dagli utenti, per preservarne il contesto originale. Per facilitare i worker nel compito di *contare* le parole, abbiamo rimosso dal testo dei tweet emoticons o caratteri particolari.

2.6 Controllo di qualità

Prima di considerare il task eseguito e restituire al worker l'*output tokens*, abbiamo attivato un controllo di qualità su una delle domande riguardanti il numero di parole che compongono il tweet.

Crowd_Frame mette a disposizione la funzione GoldCheck per effettuare questo tipo di controlli. Abbiamo aggiunto un campo nascosto in **ogni** documento dei nostri HIT: questo campo, chiamato *words*, contiene il numero di parole del tweet. In questo modo è possibile scegliere su quale tweet effettuare il GoldCheck, semplicemente aggiungendo il prefisso "GOLD-" al campo *id* del documento. La nostra funzione GoldCheck fa un confronto tra il risultato della dimensione *twitterWords* e il campo *words* del documento.

3 Amazon Mechanical Turk

3.1 Configurazione

Durante la creazione del *batch* su Amazon Mechanical Turk, abbiamo scelto di pagare *virtualmente* ogni worker \$0.01 e assegnare un solo HIT per worker, in modo da far svolgere il *batch* a 15 persone diverse.

3.2 Pubblicazione task

Come detto in precedenza, visti i limiti della piattaforma Mturk nella creazione dei task, questa viene usata **solo** per fornire al worker il link del nostro bucket S3. Oltre al link, viene fornito uno dei 15 token univoci creati in precedenza e associati ad ogni singolo HIT.

4 Conclusioni

Questa parte del progetto termina con la pubblicazione del nostro task su Mturk. Abbiamo condiviso il link dei nostri HITs sul canale Teams del corso di Social Computing e su altri canali social: i risultati verranno analizzati nella seconda parte del progetto.