



# Using annotations on Mechanical Turk to perform supervised polarity classification of Spanish customer comments



Marta R. Costa-jussà<sup>a,\*</sup>, Jens Grivolla<sup>b</sup>, Bart Mellebeek<sup>c</sup>, Francesc Benavent<sup>b</sup>, Joan Codina<sup>b</sup>, Rafael E. Banchs<sup>a</sup>

<sup>a</sup> Institute for Infocomm Research, Singapore

<sup>b</sup> Universitat Pompeu Fabra, Spain

<sup>c</sup> University of Amsterdam, The Netherlands

## ARTICLE INFO

### Article history:

Received 2 March 2011

Received in revised form 20 December 2013

Accepted 26 January 2014

Available online 17 February 2014

### Keywords:

Annotation on Mechanical Turk

HIT design

Supervised polarity classification

## ABSTRACT

One of the major bottlenecks in the development of data-driven AI Systems is the cost of reliable human annotations. The recent advent of several crowdsourcing platforms such as Amazon's Mechanical Turk, allowing requesters the access to affordable and rapid results of a global workforce, greatly facilitates the creation of massive training data. Most of the available studies on the effectiveness of crowdsourcing report on English data. We use Mechanical Turk annotations to train an Opinion Mining System to classify Spanish consumer comments. We design three different Human Intelligence Task (HIT) strategies and report high inter-annotator agreement between non-experts and expert annotators. We evaluate the advantages/drawbacks of each HIT design and show that, in our case, the use of non-expert annotations is a viable and cost-effective alternative to expert annotations.

© 2014 Elsevier Inc. All rights reserved.

## 1. Introduction

Obtaining reliable human annotations to train data-driven AI systems is often an arduous and expensive process. For this reason, crowdsourcing platforms such as Amazon's Mechanical Turk,<sup>1</sup> Crowdflower,<sup>2</sup> and others have recently attracted a lot of attention from both companies and academia. Crowdsourcing enables requesters to tap from a global pool of non-experts to obtain rapid and affordable answers to simple Human Intelligence Tasks (HITs), which can be subsequently used to train data-driven applications.

A number of recent papers on this subject point out that non-expert annotations, if produced in a sufficient quantity, can rival and even surpass the quality of expert annotations, often at a much lower cost [2,13,14]. However, this possible increase in quality depends on the task at hand and on an adequate HIT design [8]. A recent survey on this field can be found in [15].

In this paper, we evaluate the usefulness of AMT annotations to train an Opinion Mining System to detect opinionated contents (Polarity Detection) in Spanish customer comments on car brands. The main objective of this work is the comparison between expert versus non-expert annotators, rather than the analysis of the system classification of polarity. We have

\* Corresponding author.

E-mail address: [vismrc@i2r.a-star.edu.sg](mailto:vismrc@i2r.a-star.edu.sg) (M.R. Costa-jussà).

<sup>1</sup> <https://www.mturk.com>

<sup>2</sup> <http://crowdflower.com>

therefore decided to focus on a rather straightforward polarity annotation task, rather than tasks that deal with more specific difficulties encountered in opinion mining, such as dealing with ironic customer comments [12,5,1]. The implementation of the opinion mining system is also rather straightforward as it is designed to showcase the use of crowdsourced annotations, rather than advance the state of the art in opinion mining in itself [11].

Currently, a large majority of AMT tasks is designed for English speakers. Although the use of Mechanical Turk for less commonly used languages has been examined in academia (e.g. [7]), a large majority of AMT tasks is designed for English speakers. One of our reasons for using Amazon Mechanical Turk was to find out how easy it is to obtain annotated data for Spanish. This is particularly relevant as at the time of this study Mechanical Turk had only recently opened up to workers outside the US, and support for international workers remains very limited to this day. In addition, we wanted to find out how useful these data are by comparing them to expert annotations and using them as training data of an Opinion Mining System for polarity detection.

This paper is structured as follows. Section 2 contains an explanation of the task outline and our goals. Section 3 contains a description of three different HIT designs that we used in this task and Section 4 describes the competence test to check candidates' Spanish skills. In Section 5, we provide a detailed analysis of the retrieved HITs and focus on geographical information of the workers, the correlation between the different HIT designs, the quality of the retrieved answers and on the cost-effectiveness of the experiment. In Section 6, we describe the datasets that were used in the experimentation. In Section 7, we evaluate the incidence of AMT-generated annotations on a polarity classification task using two different experimental settings. Finally, we conclude in Section 8.

## 2. Task outline and goals

We compare different HIT design strategies by evaluating the usefulness of resulting Mechanical Turk (AMT) annotations to train an Opinion Mining System on Spanish consumer data. Fig. 1 shows a block diagram as schema of the task and methodology proposed. More specifically, we address the following research questions:

- (i) Annotation quality: how do the different AMT annotations compare to expert annotations?
- (ii) Annotation applicability: how does the performance of an Opinion Mining classifier vary after training on different (sub) sets of AMT and expert annotations?
- (iii) Return on Investment: how does the use of AMT annotations compare economically against the use of expert annotations?
- (iv) Language barriers [7]: currently, most AMT tasks are designed for English speakers. How easy is it to obtain reliable AMT results for Spanish?

## 3. HIT design

We selected a dataset of 1000 sentences containing user opinions on cars from the automotive section of [www.ciao.es](http://www.ciao.es) (Spanish). Ciao is a European web portal which contains user reviews and comparisons of online offers and merchants on a wide variety of products. Since 2008, Ciao is owned by Microsoft. Fig. 2 shows an example of the website.

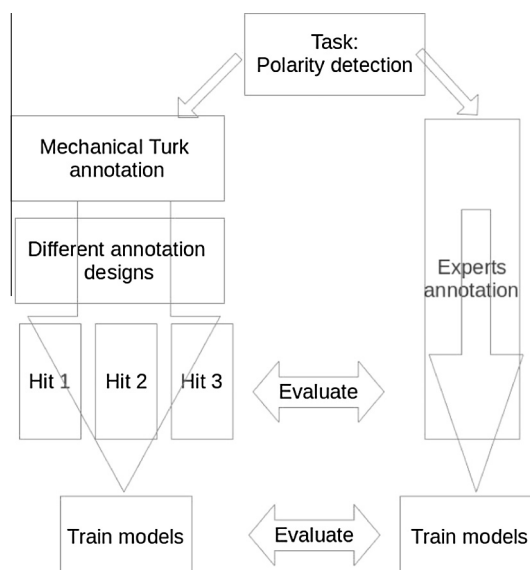


Fig. 1. Block diagram of the task.

Download English Version:

<https://daneshyari.com/en/article/392622>

Download Persian Version:

<https://daneshyari.com/article/392622>

[Daneshyari.com](https://daneshyari.com)