

Modelado de una herramienta de crowdsourcing

Emmanuel Ramírez Segura
Escuela de Computación e Informática
Universidad de Costa Rica
San José, Costa Rica
emrami@gmail.com

Juan Gabriel Gómez
Escuela de Computación e Informática
Universidad de Costa Rica
San José, Costa Rica
jgmezvargas@gmail.com

Dra. Gabriela Marín Raventós
Escuela de Computación e Informática
Universidad de Costa Rica
San José, Costa Rica
gabriela.marin@ecci.ucr.ac.cr

Msc. Edgar Casasola Murillo
Escuela de Computación e Informática
Universidad de Costa Rica
San José, Costa Rica
edgar.casasola@ecci.ucr.ac.cr

Abstract— El presente artículo aborda un modelado de una herramienta de *crowdsourcing* para ser utilizada de forma general, es decir, aplicable a cualquier temática de interés y sobre una amplia gama de aplicaciones e investigaciones. Dentro del diseño de la herramienta se mencionan aspectos fundamentales para la creación de aplicaciones de esta tipo, como lo son: el cómo mantener el interés de la audiencia, qué variables deben ser conocidas a priori en una campaña de *crowdsourcing* y se propone un modelado matemático del experimento. Finalmente, para evidenciar un caso práctico de uso de la herramienta, se plantea la creación de un corpus para análisis de sentimiento de comentarios políticos en español americano, específicamente para Costa Rica.

Keywords— *Crowdsourcing platform, sentiment analysis, human computation*

I. INTRODUCCIÓN

Crowdsourcing es un neologismo acuñado en el año 2006 por el periodista Jeff Howe para la revista *Wired*, en su artículo “The Rise of Crowdsourcing”. Él lo define como “el acto de tomar una tarea tradicionalmente realizada por un agente designado (como un empleado o un contratista) y externalizarla mediante una llamada abierta a un indefinido pero gran grupo de personas” [1].

La popularidad del *crowdsourcing* ha venido en aumento, principalmente con el uso de Internet y el desarrollo de múltiples páginas y aplicaciones del tipo WEB que incentivan a usuarios, de diferentes partes del mundo, a colaborar. Lo anterior, ha permitido optimizar los tiempos de finalización en la ejecución de tareas, así como obtener la valiosa retroalimentación de un conjunto de participantes y determinar el mayor grado de concordancia entre los mismos para obtener una respuesta “mayoritariamente más aceptada”. Su uso también ha potenciado las múltiples aplicaciones prácticas que

se le pueden dar, entre ellas, se pueden mencionar: la resolución de entidades, análisis de sentimiento, reconocimiento de imágenes[2], desambiguación del sentido de las palabras [3] y la minería de hechos [4], entre otras.

El presente trabajo se enfoca en describir de manera general, las principales ideas que se consideran deben de contemplarse detrás de la creación de una herramienta de *crowdsourcing* de uso general, es decir, aplicable a cualquier enfoque que se desee ejecutar. Paralelo a este trabajo, se encuentra en fase de desarrollo una herramienta de *crowdsourcing*, cuya intención final es que sea de uso público para que en trabajos futuros, pueda ser utilizada y evaluada.

Este artículo se estructura en siete secciones, la segunda de ellas, “Motivación en el *crowdsourcing*” propone la importancia del *crowdsourcing* y algunas sugerencias para mantener la atención del participante; así como características usadas para promover la proliferación de la evaluación por medio de los mismos votantes. La tercera sección “Generalidades de diseño del *crowdsourcing*” expone elementos de diseño que son comunes entre herramientas para este fin. La cuarta sección “Modelado Matemático” propone una descripción de las variables que intervienen en un campaña de *crowdsourcing*. La quinta sección “Uso Práctico”, brinda una descripción y una metodología para utilizar el *crowdsourcing* para crear un corpus de análisis de sentimiento con comentarios políticos. La sexta sección, contempla las conclusiones observadas y finalmente, la última sección aborda sobre posibles trabajos futuros y/o experimentos derivados del presente artículo.

II. MOTIVACIÓN EN EL CROWDSOURCING

El *crowdsourcing* es una metodología totalmente orientada hacia la gente, viendo los individuos como medio para generar información de una forma efectiva y masiva. Producto de ello,

es necesario brindarle a los participantes (votantes) una experiencia agradable mientras ejecutan la tarea asignada. Es decir, es importante lograr captar la atención del usuario, de forma que se sienta identificado y la misma vez, motivado a participar de la evaluación/votación.

Asimismo, al trabajar con una herramienta de esta clase, necesariamente se estará empleando de un modo colaborativo, lo que implica la división de una tarea, dependiendo su tamaño e índole, en microtareas o macrotareas anónimas. En el caso del análisis de sentimiento, [5] lo catalogan como una microtarea. Por otro lado, el número de herramientas crowdsourcing que circulan en el mercado es vasto, dividiendo éstas entre las que pagan a los votantes y las que, por el contrario, no son de paga. En esta última, los individuos son meramente voluntarios del experimento y no reciben remuneración económica alguna [5]. Entre las herramientas de paga se pueden mencionar algunas como MTurk, CrowdFlower [5], SurveyMonkey [6] e InnoCentive [7], mientras que algunas de las de no paga incluyen Crowd4U, Bossa, PyBossa, Zoouniverse, Crowdbutton [5], Openideo, Loby, StartSomeGood, Microplace, Sparked, Ushahidi, Causes, Change.org y Kiva [7]; lo cual deja claro que la expansión y uso de esta metodología ha tenido un gran auge durante los últimos años.

De modo que, el hecho tener la capacidad de empoderar a la gente para tomar decisiones y colaborar, ha permitido extender el número de necesidades que cubre el crowdsourcing en diferentes casos, como lo son las investigaciones científicas, manejo del tráfico vial, educación [5], interacciones entre individuos [8], política, medio ambiente, cultura, salud, arte y diseño [7], entre otros. Mientras tanto, [7] clasifica la mayoría de las razones y motivaciones que captan la atención del uso del crowdsourcing dentro de seis categorías generales, las cuales son la creatividad, decisión, recaudación de fondos, comunicación, localización y cooperación, creando así un factor común entre todas ellas y definiendo un esquema de proceso y de la razón detrás de lo que se pretende obtener.

No obstante, detrás del apoyo del evaluador, generalmente existe una razón o circunstancia que lo motiva a colaborar. Asimismo, el juez debe de tratar de percibir estas características y utilizarlas como una forma de crear un foco de atención hacia sus colaboradores, e intentando satisfacer sus expectativas; con el fin de que ambas partes obtengan resultados prometedores y las tareas sean realizadas de la mejor forma posible.

De tal manera, estudios a través de los años han logrado clasificar ciertas motivaciones del evaluador. Algunos de ellos han sido especificados por [9], quien en un contexto general categoriza el dinero, el amor y la gloria con un incentivo motivacional de la gente, el cual es aplicable al crowdsourcing. Esto, tomando también en cuenta la diversidad de herramientas y enfoques que se implementan. Igualmente, [10] agrega otros factores de motivación, donde se repite el factor dinero, pero también incluye la mejora en las habilidades, reputación en pares (aprobación, reputación, respeto), disfrute per se, autonomía de trabajo, entre otros. Esta diversidad de motivaciones permiten también enriquecer cada enfoque existente del crowdsourcing, al dar la libertad a los individuos de escoger cuál es su motivo y deseo para colaborar.

Por lo cual, tomando en cuenta los factores mencionados con anterioridad, la herramienta presentada pretende explotar el área colaborativa de la gente, ya que en la encuesta va a ser realizada utilizando una plataforma gratuita, de modo que las personas que participen deberán ser impulsadas por un ideal colaborativo.

De la misma manera, se le hará ver al colaborador la importancia de su papel en el experimento (con breves mensajes alentadores) y se incentivará el uso de la herramienta utilizando una interfaz de usuario minimalista, permitiendo al participante enfocarse de forma centrada en su tarea. Asimismo, se pretende que el votante tenga la posibilidad de compartir la encuesta a través de Facebook, con el fin de expandir el rango de usuarios que cooperan con las votaciones.

III. GENERALIDADES DE DISEÑO DEL CROWDSOURCING

En la actualidad, existe una gran variedad de herramientas de crowdsourcing, como se comentó anteriormente. A pesar de la variedad que puede haber y las diferencias de diseño entre éstas, mantienen elementos comunes que son importantes destacar y resaltar como punto de partida para la creación de nuevas aplicaciones de esta clase.

Estos elementos comunes, obedecen primordialmente al modelado matemático implícito detrás de una campaña de crowdsourcing. Por tal motivo, se expone una visión sintetizada de generalidades de diseño aplicables a la creación de una herramienta de esta índole.

Puesto que una herramienta de crowdsourcing es inherentemente una aplicación colaborativa, se identifican los siguientes actores:

El Juez. Quien puede ser una persona o grupo de personas investigadoras que tienen como meta proponer una *Tarea* a resolver.

Los Evaluadores. Quienes constituyen la audiencia o público meta de una determinada campaña de crowdsourcing. Ellos son los que ya sea de manera anónima o no, brindarán la retroalimentación necesaria para completar una *Tarea*. Quedará a criterio del *Juez* la escogencia de los evaluadores que participarán en una campaña, esto es, si se requiere que los evaluadores posean ciertos requisitos mínimos (por ejemplo, que tengan algún grado de escolaridad, que vivan en ciertas zonas geográficas, que pertenezcan a una determinada etnia, entre otros).

La Tarea. Es generalmente un cuestionario, que a través de instrucciones claras y precisas elaboradas por el *Juez*, brinda las preguntas con sus respectivas respuestas para ser escogidas por los *Evaluadores*.

El Modelo de Análisis. Este análisis toma como base la retroalimentación de las respuestas brindadas por los *Evaluadores* y a través de un algoritmo matemático estima el mayor grado de concordancia entre dichas respuestas. Algoritmos que son ampliamente utilizados y que son populares entre diversos estudios y artículos son: [11], cuya limitante más grande es que sólo obtiene grados de concordancia para dos evaluadores, y [12], el cual constituye una adaptación de [11], para ser utilizado con tres o más evaluadores. Este último es el que se utilizará como parte del diseño de la presente herramienta de crowdsourcing.

IV. MODELADO MATEMÁTICO

En esta sección se propone un modelo de las variables que intervienen cuando se realiza una campaña de crowdsourcing y sus relaciones intrínsecas.

El término campaña, se debe de entender en el presente trabajo como el envío por parte del *Juez*, de una *Tarea* hacia un grupo de *Evaluadores*, los cuales a través de su retroalimentación, permiten el análisis de la información según el grado de concordancia de las respuestas obtenidas, lo anterior, gracias al uso o aplicación de un *Modelo de Análisis*. Matemáticamente, se puede expresar una campaña *C* como la siguiente cuádrupla:

$$C = \langle J, T, E, M \rangle \quad (1)$$

Donde:

J: Juez o Jueces de la Campaña.

E: es el conjunto total de Evaluadores de una campaña *C*.

T: es la Tarea propuesta por los Jueces y que será resuelta por los Evaluadores. Para efectos del presente modelado se tratará como un cuestionario, compuesto por: P: que es el conjunto total de preguntas del cuestionario y R: que es el conjunto total de las respuestas del cuestionario.

M: El Modelo de análisis aplicado a la información.

Con estas primeras definiciones se puede partir de los siguientes hechos:

1. J propone la tarea T.
2. T puede ser visto por un conjunto P, el cual está compuesto de un subconjunto N de preguntas:

$$T = \{ P \} = \{ p_1, p_2 \dots p_N \}; N \geq 1 \quad (2)$$

3. Cada pregunta a su vez está formada de un subconjunto de respuestas diferentes (cuya cardinalidad del subconjunto de las respuestas será siempre mayor o igual a dos). Por ejemplo, la primer pregunta de un cuestionario, puede consistir de X respuestas:

$$p_1 = \{ r_{p11}, r_{p12} \dots r_{p1X} \} \quad (3)$$

Por lo tanto:

$$\text{card}(p_1) \geq 2 \quad (4)$$

Una cardinalidad mayor o igual a dos, implica poder modelar respuestas que posean múltiples opciones, donde la restricción será que el *Evaluador* seleccione sólo una de éstas.

4. Con base en los hechos anteriores, se puede relacionar la Tarea T, el Evaluador E con una nueva variable, que se definirá como Z y que corresponderá a la microtarea desarrollada por un evaluador E, matemáticamente:

$$T / E \leq Z \quad (5)$$

Donde: $T = \{ Z_0, Z_1 \dots Z_E \}$, es decir, una Tarea T estará conformada por un conjunto de microtarear Z cuya cantidad

total, será igual a la cantidad total de preguntas entre la cantidad de Evaluadores que participen de la campaña.

Por cuestiones de diseño y para no agobiar al Evaluador con una cantidad excesiva de preguntas, se aplicará la restricción de que cada Evaluador no conteste más de ocho preguntas dentro de su microtarea, por lo tanto (haciendo $Z=8$):

$$T / E \leq 8 \quad (6)$$

La fórmula anterior puede ser reescrita como:

$$E \geq T / 8 \quad (7)$$

Entonces por ejemplo, si la Tarea tiene un total de 96 preguntas, se requerirían 12 Evaluadores contestando microtarear de 8 preguntas diferentes entre sí. Pero, puesto que se requiere que una misma microtarea sea realizada al menos por tres Evaluadores diferentes (para lograr un desempate en alguna respuesta), en realidad se ocuparían, al menos, 36 Evaluadores.

Como modelo de análisis se va seguir la medición Kappa de Fleiss [12], ya que permite evaluar la concordancia entre los evaluadores de un corpus comparándolo con el azar, donde 1 es el máximo valor posible y si da menor que cero, entonces es peor que el azar. Asimismo, esta técnica “supone que hay una cantidad fija de anotaciones por instancia, pudiendo eventualmente provenir de distintos evaluadores” [13].

V. USO PRÁCTICO: CREACIÓN DE UN CORPUS DE ANALISIS SEMANTICO

Como parte de la aplicación de la herramienta descrita en este artículo, se propone la creación de un corpus de análisis semántico orientado a aspectos políticos en Costa Rica, específicamente para las elecciones del año 2018. Dicho corpus se construye con base en posibles comentarios a los cuales la multitud será responsable de asignarles información de análisis de sentimiento descritas a fondo por [14] permitiendo con ello otro insumo valioso para el análisis de la *Tarea*. Para ellos la multitud analizará los comentarios en tres formas diferentes:

Orientación Semántica. Esta categoría define como ha sido interpretado un comentario, es decir, si es un comentario es muy positivo (P+), es positivo (P), es neutro (NEU), es negativo (N), es muy negativo (N+) o ninguno de los anteriores (NONE).

Partido Político. Define si se está hablando de alguno de los siguientes tres partidos PLN (Partido Liberación Nacional), PUSC (Partido Unidad Social Cristiana) o PAC (Partido Acción Ciudadana) o Ninguno de ellos.

Aspecto. Define el tema político del cual trata el comentario, es decir, si hace referencia a un tema de Economía, Sanidad, Educación, al Propio Partido u “Otros aspectos”.

Para este caso práctico, se realizarán las siguientes consideraciones:

Los *Jueces* serán los autores del presente artículo. Los *Evaluadores* serán un conjunto de estudiantes con formación universitaria de las instituciones de educación superior costarricense, específicamente de la Universidad de Costa

Rica (UCR), Universidad Nacional (UNA) e Instituto Tecnológico de Costa Rica (ITCR). La *Tarea* será un cuestionario conformado por preguntas y respuestas elaboradas con base en comentarios extraídos (de manera aleatoria) de perfiles de usuarios costarricenses de la red social Facebook. Las respuestas relacionadas a cada comentario serán categorizadas como: *Partido Político*, *Orientación Semántica* y *Aspecto*. Como parte del *Modelo de Análisis*, se utilizará [12], con la finalidad de que no haya ninguna limitación con el número de Evaluadores que participen.

Puesto que la herramienta de crowdsourcing que se describe en este artículo, aún está en fase de diseño, se propone un ejemplo de preguntas y respuestas representativas:

Comentario Extraído Anónimamente de Red Social Facebook: “A pesar de haber heredado un país colapsado, su gestión ha sido excelente aún con la prensa corrupta desprestigiando sin tregua. Mis respetos y apoyo”

Pregunta 1. ¿El comentario puede ser relacionado con política o tema político?

Si No

Pregunta 2. Considera el comentario anterior:

Muy Positivo Positivo Neutral
 Muy Negativo Negativo Ninguno de los anteriores

Pregunta 3. Usted relaciona este comentario con el siguiente partido político:

PLN PAC PUSC Ninguno de los anteriores

Pregunta 4. ¿En cuál aspecto clasificaría mejor el comentario anterior?

Economía Sanidad Educación Propio Partido Otros

Tomando la misma idea para este comentario, se realizarían las mismas preguntas/respuestas para otros comentarios elegidos al azar. Sin embargo, por cuestiones de no hacer una microtarea muy agobiante para el Evaluador, se propone un máximo de dos comentarios por cada uno de éstos.

Los resultados obtenidos por los *Evaluadores* serán procesados por [12] como parte de un *Módulo de Análisis* dentro de la herramienta de crowdsourcing. Lo anterior permitirá obtener el corpus de análisis semántico etiquetado para comentarios políticos para Costa Rica de manera colaborativa.

VI. CONCLUSIONES

El presente artículo propone la primera descripción de modelado matemático para herramientas de Crowdsourcing, donde se define mediante una cuádrupla, las variables que intervienen a la hora de definir o realizar una Campaña (C), tales como: Juez, Evaluadores, Tarea y Modelo de Análisis, permitiendo con eso, un modelo aplicable para futuros trabajos relacionados con la creación de herramientas de esta índole. Adicionalmente, con este modelo se pretende dar la base para calcular a priori la cantidad mínima de Evaluadores requeridos para realizar una Tarea de manera distribuida.

VII. TRABAJOS FUTUROS

Con los resultados obtenidos con la herramienta de crowdsourcing, que actualmente se encuentra en fase de desarrollo, se pretende obtener un corpus etiquetado para comentarios políticos de Costa Rica, el cual a su vez, será un referente para realizar clasificaciones no supervisadas sobre otros corpus etiquetados manualmente por lingüistas y determinar mediante futuros experimentos el grado de confiabilidad de herramientas de esta clase.

VIII. REFERENCIAS

- [1] J. Howe, “Crowdsourcing Why the Power of the Crowd is Driving the future of Business,” *Human Resource Management International Digest*, vol. 18, pp. 11-16, 2010.
- [2] L. Wang, et. al. “Crowdsourced Data Management: A Survey,” *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, vol. 28, no. 7, pp. 2296-2319, 2016.
- [3] C. Akkaya, et. al. “Amazon Mechanical Turk for subjectivity word sense disambiguation,” In *Proceedings of the NAACL workshop on creating speech and language data with Amazon’s Mechanical Turk*, pp. 195–203, 2010.
- [4] R. Wang, y C. Callison-Burch. “Cheap facts and counter-facts,” In *Proceedings of the NAACL workshop on creating speech and language data with Amazon’s Mechanical Turk*, pp. 163–167, 2010.
- [5] R. Borromeo y M. Toyama, “An investigation of unpaid crowdsourcing,” *Human-centric Computing and Information Sciences*, vol. 6, no. 11, 2016.
- [6] F. Bentley, et. al. “Comparing the Reliability of Amazon Mechanical Turk and Survey Monkey to Traditional Market Research Surveys,” *Conference on Human Factors in Computing Systems Proceedings*, pp. 1092-1099, 2017.
- [7] W. Tung y G. Jordann. “Crowdsourcing social network service for social enterprise innovation,” *Springer Science Business Media*, pp. 1-17, 2017.
- [8] T. Ahmed, y A. Srivastava. “Understanding and evaluating the behavior of technical users,” *A study of developer interaction at StackOverflow*. *Hum. Cent. Comput. Inf. Sci.*, vol. 7, no. 8, 2017.
- [9] T. Malone. et. al. “Harnessing Crowds: Mapping the Genome of Collective Intelligence,” *Center for Collective Intelligence Massachusetts Institute of Technology*, no. 4732-09, 2009.
- [10] Ye, H. y A. Kankanhalli. “Solvers’ participation in crowdsourcing platforms: Examining the impacts of trust, and benefit and cost factors,” *Journal of Strategic Information Systems*, vol. 26, 2017.
- [11] J. Cohen. “A coefficient of agreement for nominal scales,” *Educational and Psychological Measurement*, vol. 20, pp. pp. 37-46, 1960.
- [12] J. Fleiss. “Measuring nominal scale agreement among many raters,” *Psychological Bulletin*, vol. 76, pp. 378-382, 1971.
- [13] S. Castro y M. Cubero. (2015). Detección de humor en textos en español. [En línea]. Disponible en: https://www.fing.edu.uy/inco/grupos/pln/prygrado/Informe_pghumor.pdf.
- [14] B. Liu. “Sentiment Analysis and Opinion Mining,” *Graeme Hirst*, 2012.