

## **USANDO AMAZON MECHANICAL TURK PARA EVALUAR IMÁGENES**

En un mundo en el que cada vez tienen lugar más relaciones y transacciones a través de aplicaciones informáticas (compras, comunicaciones y relaciones personales), muchas decisiones se toman en función del valor estético, el atractivo o el impacto de imágenes digitales. Sin embargo, existen todavía pocos resultados sobre modelización estética; es decir, sobre la capacidad de realizar valoraciones y juicios estéticos.

Diferentes grupos de investigadores han intentado crear sistemas informáticos capaces de aprender la percepción estética de un grupo de seres humanos como parte de un sistema generativo, con la intención de ser utilizados en la selección u ordenación automática de imágenes. Dada la naturaleza subjetiva del problema estético, la selección del conjunto de datos con el que se entrena el sistema es especialmente relevante.

Tras haber analizado el grado de generalización de algunos conjuntos muestrales utilizados para la clasificación automática de imágenes, se ha concluido que no es suficiente para tomarlos como referencia en el entrenamiento de sistemas de predicción y clasificación automática de imágenes. Además, se han detectado también otras limitaciones funcionales en dichos datasets.

Con la intención de ofrecer una solución a los problemas detectados, se presenta la creación de un nuevo conjunto de imágenes procedentes del portal web DPChallenge.com, con mayor coherencia estadística. Este nuevo dataset ha sido evaluado según criterios de estética y de calidad por un grupo de individuos españoles en condiciones experimentales controladas y por otro grupo de estadounidenses a través de encuestas online. Así, este dataset se convierte en el primer conjunto de imágenes evaluado por tres poblaciones diferentes (el que evalúa en el portal web DPChallenge.com, el español y el estadounidense).

Se presenta entonces la cuestión sobre el método a utilizar para la realización de las evaluaciones. Por una parte, un experimento presencial proporciona un control mucho mayor sobre las condiciones de la evaluación, pero supone un gran gasto y complica la tarea de recopilar evaluadores de un target específico. Por otra parte, un experimento vía online facilita la evaluación por parte de un gran número de personas, pero elimina el control sobre la evaluación.

Ante esta situación, es necesario elegir el método que mejor se adapte a cada ocasión.

Una herramienta muy útil para la realización de experimentos con encuestas online, es Amazon Mechanical Turk. Se trata de una plataforma crowdsourcing que requiere inteligencia humana. En dicha plataforma existen dos perfiles: los solicitantes y los trabajadores. Los solicitantes publican las encuestas con la información que desean testear y adjudican un precio de recompensa a cada tarea de la encuesta; los trabajadores realizan las tareas publicadas contestando las preguntas de las encuestas y reciben la recompensa correspondiente.