

GEAnn - Games for Engaging Annotations

Carsten Eickhoff
TU Delft
Netherlands
c.eickhoff@tudelft.nl

Christopher G. Harris
The University of Iowa
USA
christopher-
harris@uiowa.edu

Padmini Srinivasan
The University of Iowa
USA
padmini-
srinivasan@uiowa.edu

Arjen P. de Vries
CWI
Netherlands
arjen@acm.org

ABSTRACT

Crowdsourcing becomes a market of steadily growing importance on which both academia and industry, rely increasingly heavily. However, this market appears to be inherently infested with a significant share of malicious workers who try to maximise their profits through cheating or sloppiness. This serves to undermine the very merits crowdsourcing has come to represent. Based on previous experience as well as psychological insights, we propose the use of a game in crowdsourcing scenarios in order to attract and retain a larger share of entertainment seekers to relevance assessment tasks.

Motivation

In the course of the past 5 years, crowdsourcing has advanced from a niche phenomenon to becoming an accepted solution to a wide range of data acquisition challenges [1]. It has been used in the training and test phases of a great number of scientific projects and is firmly integrated into numerous evaluation cycles in academia and industry. One problem, however, seems to be inherent to the field, a significant share of the annotations created on crowdsourcing platforms are fraudsters' attempts to cheat the HIT (Human Intelligence Task) provider into paying them without having properly worked on the HIT. We hypothesize that there are 2 major types of workers with fundamentally different motivations for offering their work on a crowdsourcing platform: (1) Money-driven workers are motivated by the financial reward that the HIT promises. (2) Entertainment-driven workers primarily seek diversion but readily accept the payments as an additional incentive. We are convinced that the affiliation (or proximity) to one of those fundamental worker types can have a significant impact on the amount of attention paid to the task at hand and, subsequently, on the resulting annotations. We realize, that money-driven workers are by no means bound to deliver bad quality; however, they appear to be frequently tempted into sloppiness by the prospect of a higher time efficiency and therefore stronger satisfaction of their main motivation. Entertainment-driven workers, on the other hand, appear to do a HIT more faithfully and thoroughly and regard the financial reward as a nice bonus. They typically do not indulge in simple, repetitive or boring tasks. We propose to more strongly focus on entertainment-driven workers by phrasing crowdsourcing problems in an entertaining and engaging way: As games. We intend to increase the degree of satisfaction entertainment-driven workers experience. This, we suspect, can lead to (a) higher result quality, (b) quicker HIT uptake, (c) lower overall cheater rates. An

additional incentive for delivering high quality results in a game scenario would be the element of competition among players. Taking into account recent behavioural analyses of on-line communities and games, entertainment seekers can be expected to put much dedication into producing high quality results to earn more points in a game, progress into higher difficulty levels or a rank on the highscore.

State of the Art

Currently, the majority of crowdsourced tasks are plain surveys, relevance assessments or data collection assignments that require human intelligence but very little creativity or skill. An advance into bringing together the communities of on-line games and crowdsourcing is being made by the platform Gambit, that lets players fulfil HITs in exchange for virtual currency in their on-line game world. This combination, however, does not change the nature of the actual HIT carried out, beyond the fact that it is embedded into a game environment. With GEAnn, we propose a way to breaking up well-known query/document relevance decisions into a more engaging phrase-based association game.

Game Design

GEAnn breaks with the conventional work flow of relevance assessment, in which the judges typically view a query/document pair and determine the page's relevance towards that query. In order to make this task more entertaining, we extract keywords and images from the documents and let the players associate those to a number of topics (the original queries). Depending on their agreement with their peers they are awarded points for their decisions. In a post-processing step, we aggregate all query/keyword judgements in order to determine to which degree the content of the page is relevant to the query. As an additional benefit, this method allows us to compute passage or sentence-based relevance at no additional cost of human labour.

Future Goals

GEAnn is available on-line at www.geann.org. It is a registered participant of the TREC 2011 Crowdsourcing track, for which we aim to show the potential benefits of entertaining task design even for well-known and straightforward tasks.

1. REFERENCES

- [1] V.R. Carvalho, M. Lease, and E. Yilmaz. Crowdsourcing for search evaluation. In *ACM SIGIR Forum 2011/2*. ACM.