

Modélisation de l'incertitude et de l'imprécision de données de crowdsourcing : MONITOR

Constance Thierry*, Jean-Christophe Dubois*, Yolande Le Gall*, Arnaud Martin*

*Université de Rennes 1, 2 rue du Thabor - CS 46510, 35065 Rennes CEDEX
prenom.nom@irisa.fr,
<http://www-druid.irisa.fr>

Résumé. Le *crowdsourcing* se définit par l'externalisation de tâches à une foule de contributeurs. La foule est très diversifiée sur ces plateformes et inclut des contributeurs malveillants attirés par la rémunération des tâches et ne les réalisant pas consciencieusement. Il est essentiel d'identifier ces contributeurs afin de ne pas considérer leurs réponses. Tous les contributeurs n'ayant pas les mêmes aptitudes pour une tâche, il semble pertinent d'accorder un poids à leurs réponses en fonction de leurs qualifications. Ce papier, publié lors de la conférence ICTAI 2019, propose une méthode, MONITOR, pour l'estimation du profil du contributeur et l'agrégation des réponses par la théorie des fonctions de croyance.

1 Introduction

Le terme *crowdsourcing* définit l'externalisation de tâches à une foule de contributeurs. Dans les plateformes de *crowdsourcing*, les contributeurs réalisent des micro-tâches pour une faible rémunération. L'étude démographique de Kazai et al. (2012) montre que la foule dans ces plateformes est importante et diversifiée, introduisant une diversité dans les profils des contributeurs. Généralement, ceux-ci sont attentifs à la tâche, mais il existe néanmoins des contributeurs attirés par l'aspect financier, répondant aléatoirement et rapidement que nous nommerons *spammers*. Compte tenu de cette diversité de profils, il est difficile d'obtenir des données fiables et pertinentes issues des plateformes de *crowdsourcing*. De plus, les tâches confiées sont souvent des questionnaires à choix multiples pour lesquels le contributeur doit renseigner une réponse précise parmi l'ensemble de possibilités proposées. Or, cette interface est contraignante, car en cas d'hésitation entre plusieurs réponses, le contributeur doit en choisir une, ce qui peut l'amener à répondre aléatoirement, perdre confiance en ses capacités et générer des données peu fiables.

La méthode communément utilisée dans les plateformes de *crowdsourcing* pour l'agrégation des réponses est le Vote Majoritaire (MV) qui consiste à sélectionner la réponse renseignée par la majorité des contributeurs. Cependant, cette méthode ne considère pas l'incertitude possible sur la réponse et n'est donc pas robuste face aux *spammers*. Une alternative est l'utilisation de l'algorithme Expectation-Maximisation (EM) comme le font Raykar et Yu (2012) afin d'estimer la sensibilité et la spécificité des contributeurs, et agréger les réponses. Dans cette étude, les auteurs s'accordent sur le fait que EM est plus performant que le MV, néanmoins,