Modélisation de l'imprécision et de l'incertitude de données dans les plateformes de crowdsourcing

Constance Thierry

IUT de Lannion
Rue Edouard Branly BP 30219
22302 Lannion CEDEX, France

ENSSAT LANNION
6 rue de Kerampont
22300 Lannnion

Rapport de projet de fin d'étude

Enssat Promotion 2018 - Spécialité Informatique

Modélisation de l'imprécision et de l'incertitude de données dans les plateformes de *crowdsourcing*

Tuteurs Entreprise:
Jean-Christophe Dubois
Yolande Le Gall
Arnaud Martin

Auteur:
Constance Thierry

Tuteurs Enssat:
François Goasdoue
Daniel Rocacher

Période du stage : du 05/03/2018 au 31/08/2018
Date de soutenance : le 03/07/2018

Résumé

Le *crowdsourcing* consiste à l'externalisation de tâches à une foule de contributeurs rémunérés pour les effectuer. Il permet aux entreprises d'obtenir rapidement des résultats à bas coût. La foule, généralement très diversifiée, peut inclure des contributeurs non-qualifiés pour la tâche et/ou non-sérieux. Une bonne modélisation des réponses et de l'expertise des contributeurs est nécessaire pour exploiter au mieux les données issues des plateformes de *crowdsourcing*. Différentes méthode existent à l'heure actuelle pour modéliser les données et/ou le comportement du contributeur : utilisation d'un corpus de référence, apprentissage automatique, méthode par vote majoritaire ou probabilistes. Néanmoins ces méthodes ont leurs limites, nous nous intéressons dans ce rapport à l'utilisation des fonctions de croyances plus pertinente à notre sens. Nous présentons ici une nouvelle méthode de modélisation des réponses et de l'expertise du contributeur dans les plateformes de *crowdsourcing* se fondant sur la théorie des fonctions de croyance.

Abstract

The crowdsourcing consist to the externalization of tasks to a crowd of workers who are payed to solved them. It's allow society to obtain result quickly and few expensive. The crowd, ofen much diversify, can inclued workers who are not qualified for the task and/or non-serious. A good answer modelisation is essential exploit at te best the data comming from crowdsourcing plateformes. Nowadays existing numerous methode to modelise the data and/or the worker comportement : Gold Data uses, Machin Learning, Majority Voting, probabilities. Nethertheless these methodes have their limits. We are interested, in this report, by the used of belief functions for the modelisation of uncertainity and imprecision. We introduce a new modelisation of the worker answer and his expertise based on the belief function theory.

Table des matières

Table des figures

Liste des tableaux

Remerciement

Je tiens à remercier toutes les personnes qui ont contribué au succès de mon stage et qui m'ont aidé dans la rédaction de ce rapport.

Mes remerciements vont en premier lieu à mes encadrants de stage, Jean-Christophe Dubois, Yolande Le Gall et Arnaud Martin pour leur soutien et leurs conseils qui m'ont permis de progresser. Je leur suis reconnaissante de m'avoir offert l'opportunité de réaliser se stage de recherche au sein de leur équipe. Je remercie par ailleurs l'ensemble des membres de l'équipe DRUID de l'IRISA pour leur accueil et leur esprit d'équipe.

Enfin, je tiens à remercier mes tuteurs de l'ENSSAT, Monsieur François Goasdoue et Monsieur Daniel Rocacher pour avoir supervisé mon stage pour l'école et pour la relecture de ce rapport.

1 Introduction

De nos jours certaines tâches ne sont toujours pas réalisables par ordinateur, ou bien lorsqu'elles le sont, la réalisation n'est pas efficace et/ou très coûteuse en temps. Une solution proposée pour palier ces problèmes est l'utilisation de plateformes de *crowdsourcing*. Celles-ci permettent d'externaliser une tâche et de la répartir entre différents contributeurs. Les tâches proposées dans les plateformes de *crowdsourcing* proviennent généralement d'entreprises qui souhaitent voir cette tâche effectuée dans de brefs délais et à un coût moindre que si elle avait été réalisée par un professionnel. Le *crowdsourcing* peut être défini de manière générale comme un travail participatif qui s'accompagne d'une parallélisation des tâches.

Dans le cadre du *crowdsourcing* une tâche va être allouée à de nombreux contributeurs. Cependant la question se pose quant à la qualité des résultats des contributeurs. En effet, globalement, les plateformes de *crowdsourcing* sont ouvertes à tous, aussi les contributeurs qui souhaitent participer à une tâche n'ont pas tous le même niveau de qualification pour réaliser celle-ci. De plus, la gratification offerte pour la réalisation d'une tâche, bien qu'elle ait pour but de motiver le contributeur dans son travail, attire également des individus qui ne sont intéressés que par celle-ci. Ces contributeurs ne sont pas consciencieux dans la réalisation de leur tâche fournissant des résultats de qualité médiocre voire inutilisables. C'est pourquoi, les contributeurs ne sont gratifiés que si leurs résultats sont jugés pertinents par l'employeur ayant proposé la tâche. Ces problématiques nous amènent à la nécessité d'évaluer les résultats des contributeurs afin qu'ils puissent être exploités au mieux par l'employeur. Différentes méthodes existent à l'heure actuelle pour modéliser les données et/ou le comportement du contributeur : utilisation de données d'or, apprentissage automatique, vote majoritaire, approche probabilistes.

Les données d'or sont des corpus de référence, grâce à ces corpus l'employeur a des données de référence auxquelles il compare les réponses des contributeurs. Les articles [2, 3] font mention de l'utilisation d'apprentissage automatique sur ces données qui apportent des résultats pertinents. De manière générale, la méthode la plus employée dans les plateformes de *crowdsourcing* est la méthode par vote majoritaire qui affirme que la majorité des participants a raison [4]. Finalement des articles font état de l'utilisation d'approches probabilistes dans les plateformes de *crowdsourcing* pour la détermination de l'expertise des contributeurs et de la qualité des réponses.

L'externalisation de tâches par le *crowdsourcing* prenant de l'ampleur, les employeurs doivent s'assurer de la qualité de la réalisation de celles-ci. Comme il a été précisé ci-dessus la qualité d'une tâche dépend de l'individu ayant contribué à celle-ci. Ainsi l'objectif de mon stage réalisé dans le cadre de mes formations ENSSAT et M2SIF, au sein de l'équipe Druid de l'IRISA, est d'établir une modélisation de l'expertise du contributeur et de la qualité de sa contribution. Les méthodes existantes citées ci-dessus ont leurs limites, afin de palier à celles-ci nous utiliserons la théorie des fonctions de croyance qui s'avère être plus intéressante dans le cadre de notre travail.

Nous commençons ce rapport par la présentation de l'IRISA et de l'équipe Druid dans la section 2. Puis nous abordons dans la section 3 les défis et problématiques associés au *crowdsourcing*. Nous réalisons dans la section 4 un état de l'art afin de concevoir une cartographie des modélisations existantes pouvant répondre aux problématiques du *crowdsourcing*. Cet état de l'art nous amène à considérer la théorie des fonctions de croyances pour proposer un modèle original explicité dans la section 5. La validation de ce modèle est abordé dans la section 6, la section 7 conclue ce rapport et avance les perspectives possibles pour la continuation de nos travaux.

2 Présentation de l'entreprise

J'ai réalisé mon projet de fin d'étude au sein de l'équipe DRUID [1] de l'IRISA [2]. Dans cette section, dans un premier temps l'IRISA est présentée, puis dans un second temps l'équipe de recherche DRUID, finalement la dernière partie de cette section est consacrée aux responsabilités sociétales de l'entreprise (RSE).

2.1 L'IRISA

Le laboratoire IRISA a été créé en 1975, il s'agit d'une UMR [3] de taille importante qui a 8 tutelles (CNRS, ENS Rennes, Inria, INSA Rennes, Institut Mines-Télécom, Université Bretagne-Sud, Université de Rennes 1, CentralSupélec). Il est localisé sur quatre sites : Rennes, Lannion, Vannes et Brest. L'IRISA concentre ses recherches sur les systèmes numériques et plus particulièrement sur l'informatique et le traitement de l'information, il est organisé en sept départements (que l'on retrouve dans l'organigramme en annexe) :

D1 - Systèmes Large Échelle (LSS)
D2 - Réseaux, Télécomunication et Services (NTS)
D3 - Architecture (AAC)
D4 - Langage et génie logiciel (LES)
D5 - Signaux et Images numériques, Robotique (DSIR)
D6 - Média et interactions (MID)
D7 - Gestion des données et de la connaissance (DKM)

Les équipes de recherches sont incluses au département a la thématique associée, chaque équipe à un chef d'équipe clairement identifié (voire deux dans certains cas). Les équipes ont un agenda de recherche qui leur est propre ainsi qu'une importante autonomie aussi bien pour les aspects scientifiques que financiers.

L'équipe de recherche DRUID est une des 6 équipes de recherche du département D7, celui-ci s'intéresse au traitement des données et plus particulièrement aux relations entre données et connaissance. Les travaux de ce département portent sur le stockage, l'interrogation et la visualisation des données massives ou complexes, ainsi que sur l'exploitation et la valorisation de ces données.

2.2 L'équipe DRUID

L'équipe DRUID a la particularité d'être bi-localisée sur Rennes et Lannion, un chef d'équipe (respectivement David Gross-Amblard et Arnaud Martin) est présent sur chaque site. Mon stage s'est déroulé sur le site de Lannion. Les recherches de l'équipe portent sur la génération d'informations et connaissances fiables à partir de données incertaines produites par interaction de nombreux agents, avec un intérêt particulier pour les questions de confidentialité des données. L'équipe s'intéresse notamment aux interactions d'agents dans les réseaux sociaux et les plateformes de *crowdsourcing*.

1. Declarative & Reliable management of Uncertain, user-generated Interlinked Data
2. Institut de Recherche en Informatique et Systèmes Aléatoires
3. Unité Mixte de Recherche

Les objectifs définis par l'équipe sont les suivant :

— La coordination des utilisateurs et des tâches dans les plateformes de *crowdsourcing*.

— Le développement de théories pour la qualification des données et sources en termes de fiabilité, certitude, confiance...
Le sujet de mon stage porte principalement sur cet objectif puisqu'il s'agit de la modélisation de l'imprécision et de l'incertitude de données dans les plateformes de *crowdsourcing*par la théorie des fonctions de croyance.

— La mise en oeuvre de systèmes qui sont des preuves de concepts de ces modèles et théories
Prenons le projet HEADWORK qui est réalisé en collaboration avec le musée Cesco (Musée National d'Histoire Naturelle), la plateforme de *crowdsourcing* FouleFactory et les équipes de recherches : Valda (INRIA Paris), Links (Inria-Lille), Sumo (Inria-Bretagne). Celui-ci consiste en la réalisation d'une plateforme de *crowdsourcing* dans laquelle les contributeurs se verraient proposer des tâches suivant leur affinité avec le sujet de celles-ci. De plus, les tâches évolueraient suivant les réponses des utilisateurs, par exemple, si une tâche consiste en un apport d'information pour la biographie de l'actrice Natalie Portman il peut être demandé dans un premier temps le nombre de film dans lesquels elle a joué. Si le contributeur répond un nombre de filme aberrant, la plateforme va indiquer à celui-ci de se renseigner sur la filmographie de l'actrice. A l'inverse, si le nombre de films renseignés est crédible, la prochaine tâche du contributeur sera par exemple de citer les titres des dits films.

2.3 RSE de l'IRISA et implication de l'équipe

Gouvernance : Le Conseil de Laboratoire de l'IRISA aide à la décision de la direction pour toutes les questions relatives à la politique scientifique, la gestion des ressources, l'organisation et le fonctionnement de l'unité. Il a un rôle consultatif. Les réunions du Conseil se tiennent en moyenne 6 fois par an.

Chacun des 7 départements de l'IRISA traite un sujet stratégique de recherche. Parallèlement à ces sujets, le laboratoire considère des axes transversaux répondant à différents objectifs. Ces axes sont scientifiques (cybersécurité, biologie et santé, robotique et drone), écologique (environnement et écologie, green IT) ou encore sociétals (art, patrimoine et culture, transport). Ainsi un ou plusieurs de ces axes peuvent être centraux dans l'activité de recherche de l'équipe, applicatifs, ou encore porteurs d'intérêt sans être le sujet principal de recherche. Par exemple l'équipe DRUID ayant pour sujet principale de recherche l'extraction de connaissance et la protection de données est concernée par les axes transversaux suivants : cybersécurité, biologie et santé, environnement et écologie, art, patrimoine et culture, transport.

Droits de l'Homme : L'IRISA prévient la discrimination, notamment vis à vis des groupes vulnérables. Les locaux des sites sont aux normes pour l'accueil d'employés en situation de handicap. De plus, certaines tutelles du laboratoire telle que l'INRIA sont particulièrement actives dans leur politique handicap.

Le taux de féminisation au sein du laboratoire (TF [4]) est faible. En 2017, 9 équipes sur les 39 que compte l'IRISA avaient un TF inférieur à 10%, de manière symétrique 9 équipes ont un TF supérieur à 30%. Il n'y a pas d'équipe parfaitement mixte puisque ce TF est toujours inférieur à 45%. Cependant cette absence de totale mixité s'explique par un phénomène sociétal. En effet,

4. nombre de femmes / nombre total de membres de l'équipe

l'informatique et l'électronique sont perçus comme des domaines masculins et le pourcentage de femmes réalisant des études dans ces domaines est plus faible que celui des hommes. En 2017 d'énormes progrès sont faits en terme de recrutement féminin par l'IRISA, bien que cela n'impactes pas directement les statistiques. De plus, des actions ont été proposées pour sensibiliser et faciliter la prise de parole des femmes ainsi que pour accroître leur visibilité.

Relations et conditions de travail : L'IRISA favorise le dialogue aussi bien interne qu'externe. Pour le dialogue interne un site web proposant de nombreux outils (base de données logicielle, gestion de décision...) est mis à disposition des membres de l'IRISA. De plus différentes *mailling lists* existent pour faciliter les conversations. Parallèlement à cela différents séminaires sont réalisés au sein des départements ou encore ouvert à l'ensemble des membres du laboratoire.

Le laboratoire est sensible aux situations de travailleurs isolés et a inclus celles-ci dans son règlement intérieur. Aussi, la mise en situation de travailleur isolé doit être limitée aux circonstances exceptionnelles et faire l'objet d'une déclaration préalable auprès de la direction

Environnement : Le laboratoire prend à sa charge les déplacements des employés dans le cadre des missions de ceux-ci aussi bien pour les véhicules personnels que pour les transports en communs. De plus une politique de tri sélectif des déchets est mise en place au sein des différentes tutelles du laboratoire.

Loyauté des pratiques : Chacun est tenu de respecter la confidentialité des travaux et des informations qui lui sont confiées ainsi que celles échangées avec des tiers. En cas de présentation à l'extérieur d'éléments sensibles, l'autorisation du directeur de l'unité ou du responsable scientifique est requise. Les droits de propriété intellectuelle (brevet, droit d'auteur...) sont respectés par l'ensemble des membres de l'IRISA dans le cadre des activités de recherches.

Communauté et développement local : Une partie du personnel du laboratoire donne des cours dans les établissements de l'enseignement supérieur de Bretagne. De plus, certaines équipes participent à des manifestations telles que la fête de la science afin de partager leurs recherches et connaissances.

Enfin, certains projets portés par des équipes de l'IRISA ont un impact direct sur le développement local. Par exemple l'équipe DRUID collabore avec d'autres équipes de l'IRISA sur un projet [5] dont le principal objectif est de "réaliser la preuve de concept d'un outil d'aide à la décision publique, à la fois visuel, prédictif et généralisable". Cet outil permettrait d'identifier les territoires présentant les plus grandes chances d'accueillir favorablement de nouveaux projets. Par exemple, à partir d'un corpus d'études d'impact, cet outil aurait la capacité d'identifier rapidement les zones d'un territoire peu ou trop sollicitées sur les 5 dernières années par un type d'installation, et estimer la décision la plus probable pour une nouvelle installation similaire.

Après avoir présenté l'IRISA et l'équipe DRUID dans cette section, nous allons maintenant présenter le *crowdsourcing* et les problématiques qui lui sont associée dans la section 3.

5. Vers l'Intelligence artificielle territoriale : apprentissage massif pour la visualisation et la prédiction de décisions administratives

3 Les défis du *crowdsourcing*

Le *crowdsourcing* est une forme de production participative encore assez méconnue car il n'a pris de l'ampleur que récemment. C'est pourquoi pour une meilleure compréhension nous explicitons le principe de fonctionnement du *crowdsourcing* et les problématiques qui lui sont associées dans cette partie.

3.1 Principe de fonctionnement du *crowdsourcing*

Le principe général du *crowdsourcing* repose sur l'externalisation d'une tâche à un ensemble de contributeurs. L'accomplissement de la tâche à réaliser est en général ouvert à tous. Ainsi les contributeurs dans le cadre du *crowdsourcing* viennent de milieux divers et variés. De même, les tâches à traiter sur les plateformes de *crowdsourcing* sont également très diversifiées, ce qui permet de caractériser ces plateformes. Ainsi, Burger-Helmchen et Penin [5] mettent en évidence trois principaux types de *crowdsourcing*. Ces plateformes sont caractérisées par leurs tâches, la gratification associée et les contributeurs auxquels elles s'adressent :

- Le *crowdsourcing* d'activités routinières : les tâches qui sont réalisées dans le cadre de ce type de *crowdsourcing* ne nécessitent pas de qualifications particulières de la part des contributeurs et sont ouvertes à tous. Celles-ci proviennent en général du monde de l'industrie. Sur les plateformes de *crowdsourcing* d'activités routinières le contributeur est rémunéré pour la tâche effectuée. On retrouve ce *crowdsourcing* sur des plateformes telles que Amazon Mechanical Turk [6] ou plus connu en France FouleFactory [7].
- Le *crowdsourcing* de contenu : il s'agit en général d'un apport d'information de la part du contributeur. Il peut être rémunéré ou non. Par exemple, Wikipédia est une plateforme de *crowdsourcing* de contenu, les contributeurs viennent bénévolement apporter leur connaissance sur un domaine.
- Le *crowdsourcing* d'activités inventives : ce type de *crowdsourcing* diffère des deux précédents dans le sens où les contributeurs sont généralement moins nombreux et experts dans leur domaine. Les tâches à réaliser dans le cadre de ce *crowdsourcing* sont la résolution de problèmes, il s'apparente davantage à une forme de recherche et développement. Une plateforme connue de *crowdsourcing* de ce type est InnoCentive [8]. Une fois encore, le contributeur est rémunéré pour sa participation.

Dans le cadre d'activités routinières et inventives, le fonctionnement général du *crowdsourcing* est le suivant : une société a besoin qu'une tâche soit réalisée. Elle va alors faire appel à une plateforme de *crowdsourcing* afin que la tâche soit mise en ligne sur la plateforme et accessible à tous. Les utilisateurs de la plateforme (les contributeurs potentiel), vont alors voir apparaître la possibilité de contribuer à cette tâche. Une fois la tâche effectuée par différents contributeurs, la société récupérera les résultats auprès de la plateforme.

Le principe de fonctionnement est un peu différent pour le *crowdsourcing* de contenu. En général la foule fournit des informations et/ou des données sur les plateformes. Puis les entreprises (voire les particuliers) viennent y chercher des informations et/ou des données suivant leurs besoins. Il ne s'agit pas nécessairement d'une tâche attribuée à la foule. Si nous prenons par exemple Wikipédia,

6. https ://www.mturk.com
7. http ://www.foulefactory.com
8. https ://www.innocentive.com/

l'utilisateur vient s'informer selon ses besoins. Un autre exemple est celui de la plateforme iStockphoto[9] [6] sur laquelle diverses photos sont mises en ligne par les contributeurs et peuvent être achetées par des particuliers ou des sociétés à bas prix.

L'existence d'une telle diversité dans les activités de *crowdsourcing* est due à son utilisation massive par les industries de nos jours. En effet, comme Howe [6] l'explique le *crowdsourcing* a pris beaucoup d'ampleur au cours des dernières années. Les entreprises préfèrent désormais s'adresser à ces plateformes où elles trouveront de la main d'œuvre et/ou des données à coût moindre que celles de professionnels dans des délais de production plus courts. Howe [6] raconte l'expérience d'un photographe professionnel à qui une société s'était adressée pour des photos. Mais celle-ci a finalement fait le choix d'aller sur iStockphoto car sur cette plateforme il est possible de trouver des photos à partir de 1$ ce qui est un prix dérisoire comparé aux honoraires d'un photographe professionnel. Les principaux avantages du *crowdsourcing* pour l'industrie sont donc : son faible coût, la rapidité et la diversité des résultats grâce à la diversité des contributeurs composant la foule.

Après avoir explicité l'avantage du *crowdsourcing* pour l'industrie, on peut s'interroger sur son intérêt pour la foule : pourquoi le *crowdsourcing* fonctionne-t-il si bien auprès des contributeurs ? Felstiner [7] donne des éléments sur ce sujet, on retiendra notamment la liberté du contributeur dans son travail. En effet, le contributeur est libre de choisir le temps qu'il souhaite accorder à une tâche et de choisir la tâche qui l'intéresse. De plus, une gratification est offerte au contributeur afin de susciter chez lui de l'intérêt pour la tâche. En l'absence de gratification, l'intérêt du contributeur vient d'une motivation individuelle. Remarquons que dans le cas où une gratification est offerte, c'est principalement celle-ci qui intéresse le contributeur. Même s'il existe toujours chez certains la réelle volonté de contribuer à une tâche ce n'est pas le cas pour tous les contributeurs. La gratification attire parfois des contributeurs de "mauvaises intentions" qui ne s'intéressent pas à la tâche et la réalisent sans y porter attention. Ces contributeurs qui ne s'intéressent qu'à la gratification, et dont la contribution apparaît comme négative, sont appelés *spammers*. De même nous pouvons également spécifier l'existence de contributeurs qui bien que consciencieux dans leur travail n'ont pas les qualifications requises pour celle-ci, par la suite nous appellerons simplement ces contributeurs "non-compétant". A l'inverse, un expert est un contributeur qui remplit parfaitement la tâche qu'il a choisi car il a une excellente connaissance du sujet. Précisons que lorsqu'une tâche est rémunérée, la rémunération se fait en fonction de la qualité (estimée par l'entreprise) des réponses des contributeurs. Ainsi, un contributeur dont les résultats sont inexploitables pour l'entreprise (*spammer* et non-compétant) ne sera pas rémunéré.

On voit ici apparaître les principales problématiques du *crowdsourcing* qui sont : la motivation de la foule, la caractérisation de la foule, et ainsi l'identification des données pertinentes.

3.2 Principales problématiques associées au domaine

Cette section définit les trois principales problématiques rencontrées sur les plateformes de *crowdsourcing*. Le premier problème abordé ici est celui de la motivation de la foule, qui amène aux problèmes de la qualité des données et de la caractérisation de la foule.

9. https ://www.istockphoto.com/fr

3.2.1 La motivation de la foule

Parmi les diverses problématiques associées au *crowdsourcing*, la motivation de la foule [5, 8, 7], est la plus complexe à résoudre car il s'agit d'un problème social, sur lequel il est difficile d'inférer. Bien qu'il ne soit pas possible d'agir directement sur cette problématique, il est important de l'analyser car elle est à l'origine des problèmes de la qualité des données et de la caractérisation de la foule. Comme précisé précédemment, certaines plateformes n'offrent pas de gratifications, dans ce cadre là, la question de la motivation de la foule ne se pose pas puisque celle-ci accomplit une tâche bénévolement. C'est d'ailleurs généralement le cas pour l'ensemble des plateformes de *crowdsourcing* de contenu. Le problème de la motivation de la foule se pose davantage dans le cadre des *crowdsourcing* d'activités routinières et d'activités inventives.

Afin de palier à ce problème, les contributeurs se voient offrir une gratification (généralement une rémunération) pour leur travail. Une nouvelle différence se fait alors, puisque dans les plateformes d'activités routinières l'ensemble des contributeurs consciencieux sont rémunérés pour leur tâches. Alors que dans les plateformes inventives, la résolution d'un problème posé se fait d'avantage sous la forme d'un concours, dans ce cas seuls les contributeurs "gagnant" le concours sont rémunérés. C'est pourquoi, les *spammers* sont d'avantage présents sur les plateformes d'activités routinières, puisque sur celles d'activités inventives il est plus complexe d'obtenir une rémunération. La foule sur ces plateformes se constitue alors de contributeurs consciencieux et de *spammers*. Les contributions apportées par ces deux catégories d'individus étant très différentes en terme de qualité, il est important de les différencier en vue d'une meilleure exploitation des résultats par l'employeur. Ceci nous amène à une seconde problématique importante du *crowdsourcing* : l'identification des contributions correctes et pertinentes.

3.2.2 La qualité des réponses

L'externalisation des tâches apporte des avantages à l'entreprise, mais également un inconvénient majeur : l'employeur n'a pas de maîtrise complète sur le travail des contributeurs. Aussi, bien que les contributions de certains individus de la foule sont correctes et pertinentes pour l'utilisation que souhaite en faire l'entreprise, ce n'est malheureusement pas le cas pour toutes les contributions. C'est pourquoi, il est nécessaire de trouver une mesure qui permet de qualifier la qualité des données issues de plateformes de *crowdsourcing* afin de savoir quelles données sont exploitables et lesquelles ne le sont pas. De nombreux articles portent sur ce sujet de la qualité des données dans le cs [9, 10, 1, 11, 12, 13, 2, 14] et développent différentes méthodes pour parvenir à cette identification. Ces méthodes seront abordées dans la section 4 suivante. Certaines méthodes ne portent que sur la considération des données, d'autres en revanche s'intéressent également à la participation du contributeur. Ceci nous amène à une autre problématique importante du *crowdsourcing*, la caractérisation des contributeurs.

3.2.3 La carcatérisation de la foule

Nous différencions quatre caractères chez les contributeurs :
— le non-consciencieux (*spammer*) qui ne s'intéresse pas à la tâche.
— le non-compétant est consciencieux dans la réalisation de la tâche mais possède trop de lacunes pour l'exécuter convenablement.
— le compétant est consciencieux et qualifié pour la tâche.

— l'expert est consciencieux et a d'excellentes qualifications pour la tâche puisque comme son nom l'indique il est expert du domaine.

Il est essentiel de caractériser la foule pour traiter les réponses de celle-ci en conséquence afin d'obtenir de meilleurs résultats pour l'entreprise. Néanmoins, bien que la notion de *spammer* semble évidente il n'est pas nécessairement aisé de les différencier du reste de la foule. Prenons l'exemple d'un contributeur non-qualifié pour la tâche qu'il réalise et dont les réponses sont inexploitables. Son travail risque d'être confondu avec celui d'un *spammer* et donc ne serait pas rémunéré. On observe bien dans cet exemple la difficulté de caractériser la foule. Or, il n'existe pas encore de véritable méthode prédéfinie pour caractériser la foule, et cela fait toujours l'objet de recherches [9, 1, 11, 15]. Cette caractérisation de la foule sera abordée dans la partie 4.3.

Nous ne pouvons influer sur la motivation de la foule car celle-ci relève de l'ergonomie de la tâche et de la gratification offerte. C'est pourquoi nous nous focalisons sur les deux autres problématiques majeures que sont l'identification des données pertinentes et la caractérisation de la foule dans les plateformes d'activités routinière. Nous nous intéressons dans la partie qui suit aux modélisations existantes répondant à ces problématiques et leurs limites.

4 État de l'art

Cette partie présente une cartographie des modélisations existantes pour l'identification des réponses pertinentes et la classification des contributeurs dans les plateformes de *crowdsourcing*. Nous développons dans un premier temps, section 4.1, les modélisations actuelles n'utilisant pas la théorie des fonctions de croyance. Puis dans un second temps nous introduisons cette théorie et plus précisément les propriétés que nous exploitons pour notre modélisation dans la section 4.2. Finalement nous abordons dans la section 4.3 l'utilisation actuelle faite de la théorie des fonctions de croyance dans le cadre du *crowdsourcing*.

4.1 Limites des modélisations actuelles

Nous différencions ici les modélisations utilisant un corpus de référence de celles qui en font abstraction. Les deux premières modélisations abordées ici portant sur l'utilisation de données d'or et l'apprentissage automatique, sections 4.1.1 et 4.1.2 nécessitent l'utilisation d'un corpus de référence. À l'inverse, les deux modélisations suivantes : méthode par vote section 4.1.3 et approche probabiliste section 4.1.4 ne nécessitent pas de corpus.

4.1.1 Données d'or

Actuellement la méthode la plus simple permettant d'estimer l'expertise d'un contributeur et donc le crédit à accorder à ces réponses est l'utilisation d'un corpus de référence. Ces corpus sont appelés "données d'or" dans le cadre du *crowdsourcing* et possèdent différents intérêts dans les plateformes. Avant de réaliser une tâche le contributeur doit parfois effectuer un apprentissage pour pouvoir réaliser au mieux la tâche qu'il a choisie. Un corpus peut être utilisé dans le cadre de cet apprentissage par le contributeur pour l'aider à améliorer ses compétences [14]. Plus généralement les corpus sont séquencés et insérés dans les tâches à réaliser de façon aléatoire. Ainsi parmis l'ensemble des questions auxquelles répondra le contributeur, les réponses à certaines questions seront connues. Utilisées de la sorte, les corpus permettent d'identifier les *spammers* et les contributeurs

non-qualifiés pour la tâche. Les premiers ne portant pas de véritable considération à la tâche, leurs réponses aux questions du corpus sont fausses dans l'ensemble. Pour les seconds, leur manque d'expertise se ressent dans leurs réponses inexactes aux questions du corpus de référence.

4.1.2 Apprentissage automatique

L'apprentissage automatique peut permettre de différencier les contributeurs "consciencieux" des "non-consciencieux". Les articles [2, 3] développent la contribution de l'apprentissage automatique dans le cadre du *crowdsourcing* afin de distinguer les contributeurs "non-consciencieux' du reste de la foule. Plus précisément, Halplin et Blanco [3] abordent de façon intéressante l'utilisation de corpus de référence pour réaliser cet apprentissage. D'après leur étude l'utilisation de corpus permet de classifier au mieux *spammer* et "non-*spammer*". Ils comparent deux méthodes de classification, l'une utilisant les machines à vecteurs supports [10] (SVM) et l'autre utilisant les arbres de décisions. En se reportant à leurs expérimentations et en regardant les résultats obtenus pour la classification des *spammers* on constate que dans le meilleur des cas le taux de bonne classification est de 97.92% avec l'utilisation de SVM ce qui offre de bons résultats ; même si l'on peut craindre un sur-apprentissage.

Malheureusement toutes les tâches de *crowdsourcing* ne permettent pas l'utilisation d'un corpus de référence. Prenons l'exemple d'une question à réponse ouverte, il est difficile voire impossible de donner une réponse de référence. Cependant, quand leur utilisation est possible, il est fortement recommandé de les exploiter, Penna et Reid [16] en développent l'intérêt.

4.1.3 Méthode par vote majoritaire

En l'absence de corpus de référence, la méthode traditionnellement utilisée dans les plateformes de *crowdsourcing* pour déterminer la réponse à une question par combinaison des réponses des contributeurs, est la méthode par vote majoritaire. Bien que cette méthode ne permet pas de caractériser la foule elle est souvent utilisée car facile à implanter. Les réponses des contributeurs sont modélisées par une fonction indicatrice $\widehat{r_i}$ qui représente l'indicatrice sur la i^{eme} réponse possible à la question concernée. Ainsi, pour chaque question, sur l'ensemble des réponses des participants, la réponse qui a eu la majorité des voix est sélectionnée. Prenons le cas d'une question q binaire (par exemple la réponse à la question peut être vrai où faux), soit la réponse $r_{q,c}$ d'un contributeur c à la question q. En considérant une foule de N contributeurs Raykar *et al.* [4] définissent le vote majoritaire sur la réponse par :

$$\widehat{r_i} = \begin{cases} 1 \quad \text{si} \quad \frac{1}{N}\sum_{c=1}^{N} r_{q,c} > 0.5 \\ \\ 0 \quad \text{si} \quad \frac{1}{N}\sum_{c=1}^{N} r_{q,c} < 0.5 \end{cases} \tag{1}$$

Remarquons que l'équation (1) ne présente pas le cas d'égalité. Aussi s'il n'y a pas de majorité effective pour une réponse parmi les deux proposées le choix final de la réponse d'après Raykar *et*

10. Deux classes peuvent être séparé dans un espace de dimension n par un plan, les SVM sont des classifieurs qui on pour objectif de minimiser la marge du plan dissociant ces deux classes.

al. peut-être arbitraire ou bien demandé à un expert du domaine. La limite de cette approche réside dans le fait que tous les contributeurs sont considérés de façon similaire, avec un même poids. En effet, cette méthode ne prend pas en considération la foule mais seulement les contributions de celle-ci. Aussi elle n'est pas robuste face aux *spammers* et aux contributeurs non-qualifiés. La contribution de ces contributeurs a la même importance que celle d'un expert, ce qui peut être préjudiciable. De plus, même si l'on admettait que la foule est constituée uniquement de contributeurs consciencieux, et donc qu'aucun contributeur n'a un comportement de *spammer*, cette méthode reste imparfaite car tous les contributeurs n'ont pas le même niveau de compétences pour remplir leur tâche. Aussi, si on reprend l'exemple développé par Raykar *et al.* [4] pour illustrer la principale limite de la méthode par vote majoritaire, avec dans cette foule un expert, le reste de la foule à un niveau de qualification pour accomplir la tâche moindre que cet expert. La réponse de l'expert n'aura pas davantage d'importance, et dans le pire des cas, si elle est différente de la réponse de l'ensemble du reste de la foule, celle-ci ne sera pas retenue alors même qu'elle est plus pertinente. Comme nous pouvons le voir ici le problème principal de cette méthode est l'absence de caractérisation de la foule. Une approche probabiliste de la représentation des données a l'avantage sur l'approche par vote de prendre en compte l'expertise du contributeur [1] et de considérer l'incertitude sur la réponse du contributeur.

4.1.4 Probabilité et algorithme EM

Les approches probabilistes, permettent de modéliser l'incertitude sur les réponses d'un contributeur, et ainsi de donner un degré d'importance à cette réponse pour agir en conséquence. De la sorte, les réponses de *spammers* qui seront identifiées comme fortement incertaines et ne devront pas être considérées par l'employeur, à l'inverse des réponses des contributeurs consciencieux.

L'expertise du contributeur est réalisée grâce à l'estimation de la justesse de ses résultats aux questions posées. On se doit de dissocier dans l'approche probabiliste l'incertitude sur la réponse d'un contributeur c à une question q : $P(r_{q,c})$. De l'incertitude sur la réponse issue de l'ensemble des réponses des N contributeurs à la question q : $P(r_q|r_{q,1}, ..., r_{q,N})$.

Une approche probabiliste se fondant sur les travaux de Dawid et Skene [17] est exploité par Raykar et Yu [2]. Elle consiste à estimer les valeurs de la matrice de confusion grâce à l'algorithme *Expectation-Maximisation* (EM) introduit par Demptster[18]. La justesse des résultats d'un contributeur est calculée grâce à la matrice de confusion qui lui est associée. Une foule de N contributeurs est employée pour classifier M données. Le résultat à une question q : "l'élément y_q appartient à la classe A ou B" retourne une valeur de y_q dans $\{0, 1\}$ représentant deux classes. L'ensemble des résultats de classification des contributeurs noté $\mathcal{D}$ est tel que $\mathcal{D} = \{y_q^1, ..., y_q^N\}_{q=1}^M$. L'incertitude sur $y_q = 1$ (qui peut se traduire par l'élément y_q appartient à la classe B) est calculée par la probabilité d'appartenance $p = P[y_q = 1]$. Or connaissant $\mathcal{D}$ il est possible de calculer la probabilité de $y_q = 1$ sachant $\mathcal{D}$ pour l'ensemble des M questions : $\mu_i = P[y_q = 1|y_q^1, ..., y_q^N]$ $\forall q = 1, ..., M$. De plus, la relation suivante est donnée dans l'article :

$$p = \frac{1}{M} \sum_{q=1}^{M} \mu_q \tag{2}$$

Les auteurs de l'article utilisent l'algorithme EM sur l'ensemble des classifications $\mathcal{D}$ effectuées par les contributeurs en vu d'obtenir :

— l'ensemble des μ_q
— l'ensemble des probabilités $p = P[y_q = 1]$
— les vecteurs de sensibilité α^c et spécificité β^c sur l'ensemble des contributeurs
— un vecteur propre à l'étude menée dans le cadre des recherches de l'article

L'étape de maximisation (*M-step*) de l'algorithme EM permet d'estimer les paramètres manquants, notamment en calculant les vecteurs de sensibilité α^c et spécificité β^c. L'étape d'estimation (*E-step*) permet d'estimer les variables nécessaires en recalculant les probabilités du modèle. Ces deux étapes (*E-step* et *M-step*) sont répétées jusqu'à la convergence de l'algorithme.

Ipeirotis *et al.* [19] argumentent l'utilisation d'une technique attribuant "un score de qualité" aux travailleurs grâce à la modélisation des données par les probabilités. Cette technique a pour but de cibler les *spammers* et les contributeurs jugés non-qualifiés. Les auteurs mettent aussi en évidence par l'exemple une autre classe de contributeurs, ceux qui sur des réponses binaires répondent toujours faux. Les réponses de ce type de contributeur étant toujours fausses, il est très facile de les corriger sur des réponses binaires (en prenant toujours la réponse contraire) en vue de les exploiter. Ce type de contributeur doit être différencié du *spammer* dont les réponses sont aléatoires et inexploitables. Afin d'utiliser au mieux les données issues de plateformes de *crowdsourcing*, il est important de caractériser les contributeurs. Un premier apport de cette caractérisation est qu'elle permet de différencier les contributeurs consciencieux du reste de la foule, afin de les considérer au mieux. Mais il est intéressant d'aller plus loin et de donner un degré d'expertise aux contributeurs, ce degré d'expertise nous permettrait alors de pondérer l'impact des réponses de chaque contributeur, afin d'améliorer la qualité des contributions pour l'employeur. Cette notion de degré d'expertise a été développée dans les articles [9, 1, 15, 19].

Pour conclure sur cette partie expliquant les modélisations actuelles, résumées dans le tableau 1, bien que les méthodes avec corpus soient intéressantes elles restent difficiles à généraliser dans les plateformes puisqu'il n'est pas toujours possible d'avoir un corpus de référence. En l'absence de corpus, l'approche probabiliste est plus intéressante que la méthode par vote majoritaire car elle permet de modéliser l'incertitude sur les données et de calculer la sensibilité et la spécificité des résultats de chaque contributeur. De plus nous avons vu que Ipeirotis *et al.* [19] proposent une approche probabiliste qui permet d'exprimer une forme de degré d'expertise. Néanmoins cette forme de degré d'expertise a ses limites, car bien qu'elle prenne en compte l'incertitude sur les données, elle ne tient pas compte de l'imprécision. La théorie des fonctions de croyance permet de modéliser cette imprécision en plus de l'incertitude sur les données, ce qui nous amène à présenter cette théorie dans la section 4.2 suivante.

Méthode	Avec Corpus	Sans Corpus
Non-itérative	**Données d'or** Avantage : Facile à implanter Inconvénient : Pas toujours utilisable	**Méthode par vote majoritaire** Avantage : Facile à implanter Inconvénient : Non robuste
Itérative	**Apprentissage Automatique** Avantage : Classification contributeurs Inconvénient : Pas toujours utilisable	**EM et probabilités** Avantage : Expertise du contributeur Inconvénient : Absence de l'imprécision

TABLE 1 – Modélisations actuelles

4.2 La théorie des fonctions de croyance

La théorie des fonctions de croyance a été introduite par Dempster [20] et permet de modéliser l'incertitude et l'imprécision de sources imparfaites. Elle est une généralisation des approches floues et probabilistes. Considérons $\Omega = \{\omega_1, ..., \omega_K\}$ l'ensemble des réponses possibles à une question q posée au contributeur $c = 1, ..., N$. Cet ensemble Ω est appelé cadre de discernement. Soit $X \in 2^\Omega$, ensemble des disjonctions de Ω, la masse $m^\Omega(X)$ caractérise la croyance accordée à l'élément X. Cette fonction de masse est définie sur 2^Ω et à valeur dans $[0, 1]$, elle respecte la condition de normalisation :

$$\sum_{X \in 2^\Omega} m^\Omega(X) = 1 \tag{3}$$

Avec $2^\Omega = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_1 \cup \omega_2\}, ..., \Omega\}$, où Ω représente l'ignorance et $\emptyset$ peut être vu comme l'ouverture au monde hors Ω. Ainsi, lorsque $m^\Omega(\emptyset) = 0$ nous sommes dans un monde clos (aussi appelé monde fermé).

Un élément X de 2^Ω est appelé élément focal si $m^\Omega(X) > 0$. Soit m_c^Ω la fonction de masse sur les réponses appartenant au cadre Ω, du contributeur c. On remarquera que si les seuls éléments focaux sont les ω_k, alors m_c^Ω n'est ni plus ni moins qu'une probabilité.

On appelle fonction de masse catégorique : $m_c^\Omega(X) = 1$ avec $X \in 2^\Omega$. Cette fonction signifie que le contributeur c ne croit qu'en l'élément X et rien d'autre, ce résultat peut donc être imprécis mais il est certain. On peut ainsi modéliser l'ignorance totale, si nous avons $m_c^\Omega(\Omega) = 1$; cela signifie que le contributeur c croit en absolument toutes les réponses possibles à la question, il ne parvient pas à se positionner et choisir ce qu'il estime être la bonne réponse parmi l'ensemble des possibilités.

Un autre cas particulier est la fonction de masse à support simple (notée X^α) définie par :

$$\begin{cases} m_c^\Omega(X) = \alpha \text{ avec } X \in 2^\Omega \setminus \Omega \\ \\ m_c^\Omega(\Omega) = 1 - \alpha \\ \\ m_c^\Omega(Y) = 0 \text{ avec } Y \in 2^\Omega \setminus \{X, \Omega\} \end{cases} \tag{4}$$

Cette fonction de masse traduit le fait que le contributeur c croit en partie en X mais pas davantage, ce contributeur a une connaissance incertaine et imprécise.

Dans le cadre du *crowdsourcing*, une tâche est résolue par la foule, nous sommes donc en présence de manipulation de données imparfaites (tous les contributeurs n'ont pas la même réponse à une question q) issues de sources distinctes (les N contributeurs). Pour modéliser la fusion des informations issues des contributeurs, l'opérateur de combinaison conjonctive peut-être employé :

$$m_{Conj}^\Omega(X) = (\oplus_{c=1}^N m_c^\Omega)(X) = \sum_{Y_1 \bigcap ... \bigcap Y_N = X} \prod_{c=1}^N m_c^\Omega(Y_c) \tag{5}$$

Cet opérateur permet de réduire l'imprécision sur les éléments focaux et d'accroître de la croyance sur les éléments concordants entre les sources. Il est notamment utilisé par Ben Rjab *et al.* [11] en vu d'attribuer un degré d'expertise aux contributeurs pour caractériser la foule.

L'opérateur de combinaison conjonctive de Yager [21] donné par l'équation (6) est intéressant car il permet de rester en monde clos en interprétant la masse de l'ensemble vide comme de l'ignorance. Soit $X \in 2^{\Omega}$, $X \neq \emptyset$, $X \neq \Omega$, l'opérateur est donné par :

$$\begin{cases} m_Y^{\Omega}(X) = m_{Conj}(X) \\ m_Y^{\Omega}(\Omega) = m_{Conj}(\Omega) + m_{Conj}(\emptyset) \\ m_Y^{\Omega}(\emptyset) = 0 \end{cases} \tag{6}$$

Lorsque les informations sont issues de plusieurs cadres de discernement Ω et Θ que l'on souhaite combiner, on réalise l'extension vide sur ces cadres avant la combinaison. Elle est donnée pour tout $A \subset \Omega$ par :

$$m^{\Omega\uparrow\Omega\times\Theta}(B) = \begin{cases} m^{\Omega}(A) \text{ si } B = A \times \Theta \\ 0 \text{ sinon} \end{cases} \tag{7}$$

De même, considérant le produit cartésien $\Omega \times \Theta$ on peut se projeter sur le cadre de discernement Θ (respectivement Ω) en réalisant une marginalisation sur la fonction de masse issue de la combinaison. Ainsi, une fonction de masse pour la marginalisation de $\Omega \times \Theta \downarrow \Theta$ est donnée $\forall B \subseteq \Omega$:

$$m^{\Omega\times\Theta\downarrow\Omega}(B) = \sum_{A\subseteq\Omega\times\Theta, A\downarrow\Omega=B} m^{\Omega\times\Theta}(A) \tag{8}$$

Afin de prendre une décision sur les éléments du cadre de discernement Ω nous considérons l'élément focal $\omega_i \in \Omega$ pour lequel on obtient la probabilité pignistique maximale $betP$:

$$betP(\omega_i) = \max_{\omega\in\Omega} betP(\omega) \tag{9}$$

Avec $betP$ définie par l'équation :

$$betP(X) = \sum_{Y\in 2^{\Omega}, Y\neq\emptyset} \frac{|X \cap Y|}{|Y|} \frac{m^{\Omega}(Y)}{1 - m^{\Omega}(\emptyset)} \tag{10}$$

L'utilisation de cette théorie dans le cadre du *crowdsourcing* est intéressante car elle permet de modéliser l'incertitude et l'imprécision des réponses d'un contributeur ainsi que la fusion des données issues de l'ensemble des contributeurs vus comme des sources d'information. Après avoir introduit les fonctions de croyance, nous abordons dans la section suivante son utilisation dans le cadre du *crowdsourcing*.

4.3 Fonctions de croyance et *crowdsourcing*

L'utilisation de la théorie des fonction de croyance dans les plateformes de *crowdsourcing* est intéressante car elle permet de modéliser l'incertitude et l'imprécision sur les réponses des contributeurs. Ces réponses sont incertaines puisque la réponse "juste" n'est pas connue du contributeur, et peuvent être imprécises dans le cas de réponse multiple et ou de réponse ouverte.

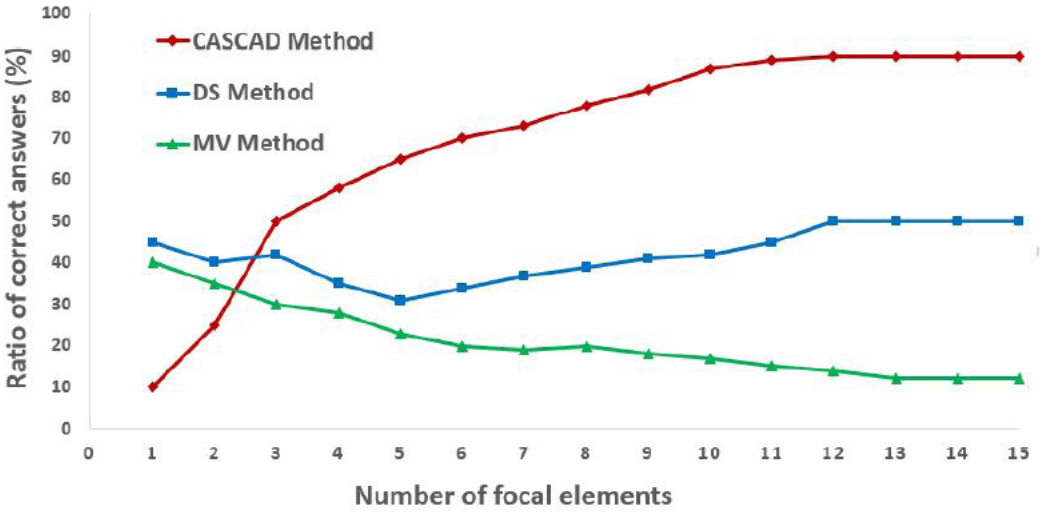


FIGURE 1 – Graphique extrait de l'article [1] : Nombre moyen de réponse correct en fonction du nombre d'éléments focaux

Nous montrons d'abord ici l'apport de la théorie des fonctions de croyance au *crowdsourcing* comparées à certaines modélisations évoquées dans la partie ci-dessus. Nous abordons dans un second temps l'utilisation des données d'or conjointement aux fonctions de croyance. Puis nous soulignons l'intérêt de modéliser l'imprécision du contributeur, et finalement nous abordons la possibilité de caractériser le contributeur par un degré d'expertise grâce aux fonctions de croyance.

4.3.1 Apport de la théorie des fonctions de croyance

Dans leur étude portant sur la combinaison des réponses dans les plateformes de *crowdsourcing*, Koulougli et *al.* [1] comparent trois méthodes fondées sur : le vote majoritaire (section 4.1.3), l'approche probabiliste de Dawide-Skene (section 4.1.4) et les fonctions de croyance. La méthode utilisant les fonctions de croyances discuté dans l'article est appelé CASCAD par les auteurs. Cette étude fait apparaître que les fonctions de croyance offrent de meilleurs résultats pour la combinaison des réponses, devant respectivement l'approche probabiliste de Dawide-Skene et le vote majoritaire. En effet, comme il est possible de le constater sur la figure 1 extraite de l'article [1], lorsque le nombre d'éléments focaux considérés pour les réponses est supérieur à 2, le ratio de bonne réponse pour la méthode CASCAD est supérieur ou égale à 50%. De plus, on constate graphiquement un écart important entre le ratio offert par la méthode CASCADE et ceux de l'approche probabiliste et du vote majoritaire. Alors que les probabilités permettent seulement la mesure de l'incertitude sur les données, les fonctions de croyance permettent également de modéliser l'imprécision, ce qui est un atout majeur dans le cadre du *crowdsourcing*.

De plus, prenons deux contributeurs c_1 et c_2, considérons l'approche probabiliste de la modélisation des données. Si les réponses de ces contributeurs ont le même degré d'incertitude alors il est normal de considérer que ces deux contributeurs ont le même degré d'expertise. Cependant, considérons maintenant la modélisation des données par la théorie des fonctions de croyance, les réponses de ces deux contributeurs ont toujours le même degré d'incertitude, en revanche, les réponses du contribu-

teur c_1 ont un niveau d'imprécision plus faible que celles du contributeur c_2. Les deux contributeurs contrairement à précédemment n'ont alors plus le même degré d'expertise, celui du contributeur c_1 étant plus élevé que celui du contributeur c_2 puisque celui-ci est moins imprécis dans ses réponses. D'où l'intérêt de l'utilisation des fonctions de croyance pour traduire une mesure d'expertise dans le *crowdsourcing* [9, 11, 22, 23]. Ces fonctions permettent une bonne modélisation de l'incertitude et de l'imprécision des données.

Comme nous avons pu le voir dans l'exemple donné ci-dessus il est intéressant de considérer incertitude et imprécision dans la caractérisation d'expert car elles sont complémentaires. La méthode de Koulougli *et al.* [1] utilisant les fonctions de croyance offre de meilleurs résultats en terme de combinaison des réponses des contributeurs que la méthode par vote et la méthode sur les probabilités (Dawid-Skene). Nous commentons dans la section suivante l'intérêt d'utiliser des données d'or conjointement aux fonctions de croyance.

4.3.2 Fonctions de croyance en présence de données d'or

Nous avons vu dans la section 4.1.2 que l'utilisation d'un corpus de référence (appelé données d'or) pouvait permettre la réalisation d'un apprentissage semi-supervisé sur les données. Ce corpus peut également être utilisé conjointement à la théorie des fonctions de croyance [9, 24]. Dans ce cas, les résultats des contributeurs peuvent être comparés au corpus afin de mieux estimer leur expertise.

Dans l'approche de Ouni *et al.* [9], un corpus de référence permet de modéliser un graphe de référence orienté, puis un graphe des réponses apportées à ces mêmes questions est construit pour chaque contributeur. Dans cet article $N_{1,q}$ représente le nœud d'attribut q (associé à la question q) dans le graphe de référence et $N_{2,q}$ est le nœud d'attribut q dans le graphe représentant les réponses d'un contributeur aux questions de corpus. L'objectif de cette approche est de mesurer l'expertise des contributeurs en vue de différencier les experts des "non-experts", c'est pourquoi le cadre de discernement est le suivant : $\Omega = \{E, NE\}$ où E signifie expert et NE non-expert. Les fonctions de croyance sont utilisées pour comparer ces deux graphes en calculant :

— Un degré d'exactitude : ce degré compare la position d'un nœud q entre les deux graphes. Ce degré est calculé en mesurant la distance Euclidienne d_1 séparant les nœuds dans les deux graphes et d_{max} la distance Euclidienne maximale antre deux nœuds sur l'ensemble du graphe.

$$\begin{cases} m_1(N_{1,q}, N_{2,q})(E) &= 1 - \dfrac{d_1(N_{1,q}, N_{2,q})}{d_{max}} \\ \\ m_1(N_{1,q}, N_{2,q})(NE) &= \dfrac{d_1(N_{1,q}, N_{2,q})}{d_{max}} \end{cases} \tag{11}$$

Ainsi si un nœud q est exactement à la place ou il devrait être dans le graphe du contributeur, la distance Euclidienne entre les deux nœuds est nulle ce qui maximise $m_1(N_{1,q}, N_{2,q})(E)$ et minimise donc $m_1(N_{1,q}, N_{2,q})(NE)$.

— Un degré de confusion : mesure la proportion de nœuds de même distance au point de départ du graphe que le nœud q. La dissimilarité de Jaccard d_2 est utilisée pour calculer la fonction

de masse associée à ce degré.

$$\begin{cases} m_2(N_{1,q}, N_{2,q})(E) &= d_2(N_{1,q}, N_{2,q}) \\ m_2(N_{1,q}, N_{2,q})(NE) &= 1 - d_2(N_{1,q}, N_{2,q}) \end{cases} \tag{12}$$

Une fois encore, lorsque la valeur de $m_2(N_{1,q}, N_{2,q})(E)$ augmente celle de $m_2(N_{1,q}, N_{2,q})(NE)$ diminue de manière évidente, ce qui est en accord avec les attentes de la différenciation d'expert et non-expert.

— Un degré de mauvais ordre précédent et suivant : ce degré mesure les erreurs d'inversion entre les nœuds précédents et suivants d'un nœud q. Pour mesurer ce degré les distances d_3 et d_4 sont calculées grâce aux définitions des prédécesseurs et successeurs données dans l'article.

$$\begin{cases} m_3(N_{1,q}, N_{2,q})(E) &= d_{3,1}(N_{1,q}, N_{2,q}) \\ m_3(N_{1,q}, N_{2,q})(NE) &= d_{3,2}(N_{1,q}, N_{2,q}) \end{cases} \tag{13}$$

$$\begin{cases} m_4(N_{1,i}, N_{2,q})(E) &= d_{4,1}(N_{1,q}, N_{2,q}) \\ m_4(N_{1,i}, N_{2,q})(NE) &= d_{4,2}(N_{1,q}, N_{2,q}) \end{cases} \tag{14}$$

L'ensemble de ces degrés permet de modéliser le degré d'expertise du contributeur c pour un degré k grâce au calcul d'une fonction de masse sur l'ensemble des graphes par la moyenne :

$$\begin{cases} m_k(G_1, G_2)(E) &= \frac{\sum_{q=1}^{O(G)} m_k(N_{1,q}, N_{2,q})(E)}{O(G)} \\ m_k(G_1, G_2)(NE) &= \frac{\sum_{q=1}^{O(G)} m_k(N_{1,q}, N_{2,q})(NE)}{O(G)} \end{cases} \tag{15}$$

$O(G)$ est le nombre de sommets du graphe. On peut noter qu'ici la fonction de masse est une probabilité puisque la possibilité d'être imprécis dans ses réponses n'est pas offerte au contributeur. Ce degré d'expertise permet d'estimer si les réponses doivent être retenues ou non par l'employeur. Encore une fois le corpus de référence permet une meilleure estimation de la contribution de la foule, ainsi il est intéressant de les utiliser conjointement à la théorie des fonctions de croyance lorsque cela est possible. Dans le cadre de l'étude menée par Ouni *et al.* [9] une campagne de *crowdsourcing* portant sur des écoutes sonores a été menée en Asie et une autre aux États-Unis. Les degrés d'expertise de l'ensemble des contributeurs des deux campagnes sont ensuite comparés, on observe une différence dans ces degrés, celui des participants de la campagne américaine étant généralement plus élevé que celui des participants de la campagne d'Asie. Les auteurs expliquent cette différence de degré par les différences culturelles entre Amérique et Asie, les premiers contributeurs étant plus familiarisés aux extraits musicaux proposés. On a ici un exemple pertinent de la caractérisation de la foule par la théorie des fonctions de croyance. On constate qu'il est intéressant d'attribuer au contributeur un degré d'expertise, car ce degré permet la comparaison entre contributeur, mais aussi entre campagne portant sur le même sujet.

Étant donnée qu'il est difficile d'avoir un corpus de référence, l'étude de Abassi et Boukhris [24] est intéressante. En effet, les auteurs utilisent aussi ce corpus, mais dans leur modélisation, celui-ci a une importance moindre que dans celle de Ouni *et al.*. De plus les auteurs s'imposent de limiter

la taille du corpus. La modélisation de Abassi et Boukhris permet de classifier les contributeurs en "Expert", "Bon contributeur" et "Mauvais contributeur", elle a pour finalité d'établir des réponses de qualité à la campagne réalisé. Cette modélisation, appelée CGS-BLA par les auteurs, peut être décomposée en trois étapes. Dans un premier temps, trois mesures de la précision des réponses du contributeur c sont réalisées en vue d'estimer à quelle classe d'expertise il appartient, pour un nombre de questions Q ces mesures sont les suivantes :

— Mesure par un corpus de référence :

$$m_{1,c} = \frac{Nb\_reponses\_corpus\_juste}{Nb\_reponses\_corpus}$$

— Mesure par MV :

$$m_{2,c} = \frac{Nb\_reponses\_juste\_par\_MV}{Q}$$

— Mesure par distance logarithmique :

$$m_{3,c} = -\frac{1}{Q}\sum_{i=1}^{Q}\ln(p(q_i))$$

Ici $p(q_i)$ est la proportion de réponses des contributeurs (autre que le contributeur c) qui sont similaires à la réponse du contributeur c.

Une fois les mesures $m_{1,c}$, $m_{2,c}$ et $m_{3,c}$ obtenus, l'algorithme de *clustering k-mean* est appliqué à ces mesures (avec k=3) afin de classifier les contributeurs. Le *cluster* dont la moyenne des mesures est la plus élevée est celui des "Experts", celui dont la moyenne est la plus faible est celui des "Mauvais contributeurs". Une fois le contributeur classifié on associe une fonction de masse à sa réponse qui est constituée d'un singleton. Ainsi la réponse d'un contributeur classifié expert est estimée certaine et une fonction de masse catégorique est associée à ces réponses. A l'inverse aucun crédit n'est attribué à un contributeur classifié "Mauvais contributeur" et ces réponses sont modélisées par l'ignorance. Au contributeur classifié "Bon contributeur", une fonction de masse à support simple est associée à ses réponses. Finalement, pour chaque question les réponses des contributeurs sont combinées, cela est possible car bien que les fonctions soient différentes il s'agit du même cadre de discernement sur les réponses. Après la combinaison la probabilité pignistique de chaque réponse est calculée afin de prendre une décision sur celle qui sera choisit. Cette méthode est intéressante car contrairement à celle de Ouni *et al.* elle ne repose pas entièrement et uniquement sur un corpus de référence. Il serait d'ailleurs possible d'imaginer la modélisation en se contentant des mesures $m_{2,c}$ et $m_{3,c}$ en l'absence de corpus.

Les deux modélisations présentées ici font usage d'un corpus de référence et considèrent des réponses précises. Or il n'est pas toujours possible d'avoir ce corpus de référence, de plus il serait également intéressant d'offrir au contributeur la possibilité d'être imprécis dans ces réponses. Ainsi la section suivante aborde l'intérêt de réponses imprécises dans le cadre du *crowdsourcing* en l'absence de corpus de référence.

4.3.3 L'imprécision du contributeur

Généralement, l'employeur exige que la foule réponde à toutes les questions, cependant certains travaux dans le domaine du *crowdsourcing* autorisent le contributeur à ne pas répondre à une

question s'il le souhaite. Cette approche est intéressante car dans le cadre de son utilisation avec la théorie des fonctions de croyance, elle permet au contributeur de ne pas se compromettre dans le cas où il ne sait pas quelle réponse donner [11]. L'absence de réponse à une question n'est pas un problème dans le cadre de l'utilisation de la théorie des fonctions de croyance grâce à la modélisation de l'ignorance, contrairement à l'utilisation de la méthode par vote ou l'absence de réponse est pénalisante. De plus avoir la possibilité de s'abstenir de répondre à une question dans le cas où le contributeur ne connaît pas la réponse permet d'éviter de considérer une réponse trop incertaine et imprécise et donc diminue l'incertitude et l'imprécision globales sur l'ensemble des réponses de la foule à cette question. On voit donc ici qu'il est intéressant d'offrir au contributeur la possibilité de ne pas répondre à certaines questions. On remarquera cependant qu'il n'est bien sûr pas indiqué à l'employeur de permettre au contributeur de passer trop de questions. En revanche il serait intéressant d'estimer la difficulté des questions de la tâche et d'autoriser aux contributeurs de passer les questions les plus difficiles.

Il est également attendu des contributeurs qu'ils soient précis dans les tâches réalisées, néanmoins, il est intéressant d'offrir au contributeur la possibilité d'être imprécis dans ses réponses tel que proposé par Ben Rjab *et al.* [11]. Ceux-ci s'intéressent à l'identification d'experts par la modélisation d'un degré de précision DP_c (17) et d'un degré d'exactitude DE_c (16) sur la réponse d'un contributeur. Soit E_c l'ensemble des contributeurs, E_{c_Q} l'ensemble des questions auxquelles un contributeur c a répondu et Ω_q le cadre de discernement associé à la question q :

$$\begin{cases} DE_c = 1 - \dfrac{1}{|E_{c_Q}|} \displaystyle\sum_{q \in E_{c_Q}} d_J(m_c^{\Omega_q}, m_{E_c|c}^{\Omega_q}) \\ \\ m_{E_c|c}^{\Omega_q}(X) = \dfrac{1}{|E_c| - 1} \displaystyle\sum_{j \in E_c|c} m_j^{\Omega_q}(X) \end{cases} \tag{16}$$

$$\begin{cases} DP_c = \dfrac{1}{|E_{c_Q}|} \displaystyle\sum_{q \in E_{c_Q}} \delta_c^{\Omega_q} \\ \\ \delta_c^{\Omega_q} = 1 - \displaystyle\sum_{X \in 2^{\Omega_q}} m_c^{\Omega_q}(X) \frac{log_2(|X|)}{log_2(|\Omega_q|)} \end{cases} \tag{17}$$

Dans l'équation (16) d_J est la distance de Jousselme [25] entre la masse $m_c^{\Omega_q}$ et la moyenne des masses $m_{E_c|c}^{\Omega_q}$. Un degré global d'expertise du contributeur GD_c est ensuite calculé en pondérant les degrés DE_c et DP_c par un coefficient $\beta_c \in [0, 1]$:

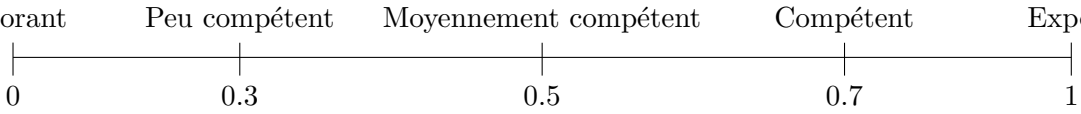
$$DG_c = \beta_c DE_c + (1 - \beta_c) DP_c \tag{18}$$

L'étude de Ben Rjab *et al.* [11] propose une comparaison avec une approche probabiliste mesurant l'expertise d'un contributeur. Leur approche reposant sur le degré GD_c s'est avérée être plus performante pour l'évaluation d'experts que l'approche probabiliste.

Nous avons constaté que les deux modélisations explicitées ici ([9] et [11]) utilisant les fonctions de croyance avec utilisation ou non de données d'or mesurent un degré d'expertise pour caractériser le contributeur. C'est pourquoi dans la section suivante nous nous intéressons à cette caractérisation par un degré d'expertise, considérant notamment la classification du contributeur suivant son degré.

4.3.4 Un degré d'expertise pour caractériser le contributeur

Une fois le degré d'expertise calculé, il est possible de l'échelonner et/ou de classifier les contributeurs en différents groupes suivant ce degré, Koulougli *et al.* [1] utilisent l'échelle de degré d'expertise suivante :



En détaillant cette échelle suivant l'article nous avons :

- l'*ignorant* est contributeur dont les résultats sont faux dans l'ensemble et inexploitables.
- le *peu compétent* n'a que très peu d'expertise pour remplir sa tâche, la plupart du temps ses réponses sont fausses.
- le *moyennement compétent* a les connaissances générales pour réaliser sa tâche, il n'est pas brillant mais fait peu d'erreurs et ses résultats sont exploitables.
- le *compétent* a une bonne connaissance du domaine dans lequel il effectue sa tâche et ses réponses sont pertinentes, l'employeur peut avoir une bonne confiance en ses résultats.
- l'*expert* est le contributeur idéal pour l'employeur, il a une maîtrise parfaite du sujet et ses réponses ont un niveau de précision et certitude élevées.

Cette échelle apporte une caractérisation intéressantes de l'utilisateur. L'employeur peut sélectionner ainsi une ou plusieurs classes de contributeurs dont il souhait exploiter les résultats suivant la flexibilité qu'il est près à accorder sur la pertinence des réponses.

De façon plus générale, l'utilisation d'un degré d'expertise permet de sélectionner l'ensemble des contributions fournies par des contributeurs dont le degré d'expertise est supérieur ou égal à un degré de seuil [9, 11, 1]. Ainsi grâce à ce degré d'expertise l'employeur peut identifier rapidement et efficacement les individus dont la contribution apparaîtrait comme non-pertinente. De plus, comme il a été vu précédemment [9], dans le cadre de campagnes réalisées à différentes reprises dans différents pays il est possible de comparer les campagnes grâce aux degrés d'expertise des contributeurs. Dans l'ensemble, le degré d'expertise permet de comparer les contributeurs alors que ceux-ci ont travaillé dans des environnements et des conditions différentes et ont des qualifications pour des tâches variées. De manière plus général, l'employeur a grâce à ce degré d'expertise une bonne caractérisation de la foule qui a contribué à la tâche qu'il a proposée.

Dans cette section 4.3 portant sur l'utilisation des fonctions de croyance pour la caractérisation des contributeurs et de leurs contributions deux approches ressortent principalement (rappelées dans le tableau 2). Une première forme d'approche nécessite un corpus de référence, l'avantage de celui-ci est qu'il permet une bonne estimation de l'expertise du contributeur pour la tâche, malheureusement il est parfois complexe voire impossible d'avoir ce corpus. Les modélisations rencontrées dans cet état de l'art en présence de corpus de référence considèrent ici des réponses précises (singleton). A l'inverse les modélisations n'utilisant pas de corpus offrent la possibilité au contributeur d'être imprécis. L'absence de corpus rend plus difficile l'estimation de l'expertise du contributeur, néanmoins offrir à celui-ci la possibilité d'être imprécis permet d'accroître la certitude en ses réponses et ainsi de parvenir d'amoindrir le biais sur l'estimation de son expertise.

Avec Corpus Sans imprécision	Sans Corpus Avec Imprécision
Distance entre graphes Ouni *et al.* [9]	**Mesure d'exactitude et de précision** Ben Rjab *et al.* [11]
Apprentissage automatique Abassi et Boukhris [24]	**Pondération par degré d'expertise** Koulougli *et al.* [1]

TABLE 2 – Modélisations utilisant les fonctions de croyance

L'état de l'art réalisé ici nous permet d'aborder avec des connaissances pertinentes la modélisation que nous proposons pour la caractérisation de la foule dans les plateformes de *crowdsourcing* en vue d'une meilleure exploitation des résultats.

5 Caractérisation des contributions et des contributeurs

Les résultats obtenus en terme de caractérisation des contributions et des contributeurs, dans les plateformes de *crowdsourcing*, en présence d'un corpus de référence sont généralement bons. Néanmoins il n'est pas toujours possible d'avoir ces corpus, c'est pourquoi nous nous intéressons à la modélisation de la qualité des contributions et de l'expertise des contributeurs en l'absence de corpus de référence. Cette modélisation présente ainsi l'intérêt de pouvoir être appliquée à l'ensemble des questionnaires proposés sur les plateformes de *crowdsourcing*. Il a été constaté dans l'état de l'art qu'en l'absence de corpus de référence il est bon de laisser au contributeur la possibilité d'être imprécis dans ses réponses, aussi cette imprécision est prise en compte dans notre modélisation.

Cette section présente l'approche que nous proposons, reposant sur la théorie des fonctions de croyance. Dans un premier temps, la modélisation de la réponse du contributeur est explicitée. Puis dans un second temps, la modélisation de l'expertise du contributeur considérant la connaissance et le comportement de celui-ci est abordée.

5.1 Modélisation de la réponse du contributeur

Dans les plateformes de *crowdsourcing*d'activités routinières, les tâches consistent généralement en des questions posées sous forme de questionnaires à choix multiple. Cette forme de questionnaire est pris en considération ici pour la modélisation des contributions. De plus l'hypothèse est faite que lorsque l'utilisateur répond à une question il renseigne également la confiance qu'il a en sa réponse.

Pour une question q considérant les réponses à cette question $\omega_1, ..., \omega_K$, nous définissons alors le cadre de discernement sur les réponses : $\Omega_1 = \{\omega_1, ..., \omega_K\}$. Il n'est pas possible pour le contributeur de renseigner une réponse autre que celles proposées par le questionnaire, aussi cette modélisation se place t'elle en politique de monde clos. Soit une réponse $X \in 2^{\Omega_1}$ d'un contributeur c à une question q, nous modélisons cette réponse par une fonction de masse à support simple (notée $X^{\alpha_{cq}}$) : $m^{\Omega_1}_{c\_q}$. La masse α_{cq} est la valeur numérique de l'incertitude de l'utilisateur sur sa réponse.

Cette modélisation de la réponse du contributeur va nous permettre de définir par la suite une mesure de l'exactitude des réponses de l'utilisateur en vue d'extraire une connaissance sur celui-ci.

5.2 Modélisation du profile du contributeur

Nous modélisons ici l'expertise du contributeur. Dans cet objectif, nous considérons sa connaissance sur le sujet de la tâche, et sa motivation à travers l'étude de son comportement dans la réalisation de celle-ci.

5.2.1 Connaissance

Nous considérons que la connaissance de l'utilisateur repose sur sa Qualification pour la tâche. Nous définissons la Qualification comme une appréciation de la valeur professionnelle d'un contributeur en fonction de l'exactitude des réponses qu'il fournit aux questions posées. Afin d'établir une mesure de l'exactitude des réponses des contributeurs sans utiliser de corpus de référence, nous exploitons certains concepts de Ben Rjab *et al.* [11].

Nous considérons le fait qu'un individu soit qualifié "Q" ou non "NQ" pour une tâche et le cadre de discernement : $\Omega_2 = \{Q, NQ\}$. Comme nous nous reposons sur l'exactitude des réponses du contributeur pour estimer sa Qualification, nous associons à la masse $m_c^{\Omega_2}$ le degré d'exactitude de DE_c défini par Ben Rjab *et al.* [11]. La fonction de masse sur la réponse de l'utilisateur est donnée par l'équation (16) : $m_c^{\Omega_q} = m_{c\_q}^{\Omega_1}$. Afin de rester en monde clos nous associons une masse nulle à l'élément $\emptyset \in 2^{\Omega_2}$. Nous associons DE_c à l'élément focal "Q" et $1 - DE_c$ à "NQ". Pour ne pas être catégorique sur la Qualification du contributeur, nous interprétons l'imprécision sur ses réponses aux questions comme de l'ignorance. Nous affaiblissons ainsi l'information donnée par le contributeur avec le degré de précision DP_c donné par l'équation (17) où $\Omega_q = \Omega_1$. Nous obtenons alors la fonction de masse :

$$\begin{cases} m_c^{\Omega_2}(Q) = DP_c * DE_c \\ m_c^{\Omega_2}(NQ) = DP_c * (1 - DE_c) \\ m_c^{\Omega_2}(\Omega_2) = 1 - DP_c \end{cases} \tag{19}$$

La modélisation proposée de la qualification du contributeur ainsi définie prend en considération non seulement l'exactitude de ses réponses, mais également l'imprécision de celles-ci.

5.2.2 Comportement

Le comportement du contributeur induit ici sa motivation dans la réalisation de la tâche considérée. C'est pourquoi nous nous sommes intéressés à la "conscience" d'un contributeur qui se traduit par son implication dans la réalisation de sa tâche. La conscience est l'une des cinq caractéristiques définies dans le modèle des *Big Five*, aussi appelé modèle OCEAN, proposé par Goldberg [26] pour caractériser la personnalité d'un individu. Les cinq caractéristiques sont l'Ouverture à l'expérience, la Conscience, l'Extraversion, l'Agréabilité et le Névrosisme. Dans une précédente étude, Kazai et al. [15] qui ont introduit ce modèle dans le contexte du *crowdsourcing* pour déterminer la relation entre les traits de personnalité et la qualité des réponses, ont conclu que la Conscience a un impact fort sur l'exactitude des résultats du contributeur. Dans notre approche, nous estimons la Conscience du contributeur en considérant sa Réflexion et utilisons à cet effet le temps de réponse du contributeur à une question qui est supposé connu.

Nous nous intéressons au temps de réflexion pris par le contributeur pour donner sa réponse, aussi nous considérons le cadre de discernement suivant : $\Omega_3 = \{R, NR\}$. "R" signifie que la réponse

du contributeur est réfléchie et "NR" qu'elle est instinctive. Soit un élément $X \in 2^{\Omega_3}$ indiquant la Réflexion du contributeur pour une question q, nous définissons la fonction de masse associée par :

$$m_{c_q}^{\Omega_3}(X) = g(T_{c_q}, T_{th_q}, X) \tag{20}$$

La fonction g est définie en annexe à la page 36 par le pseudo-code de l'algorithme 1, T_{c_q} est le temps de réponse du contributeur c à la question q et T_{th_q} un temps de réponse théorique attendue. Dans cette fonction g, la réflexion du contributeur est initialisée à instinctive (NR).

Dans ce pseudo-code, la fonction $alpha(T_{c_q}, T_{th_q})$ retourne une valeur entre 0 et 1, elle est définie elle aussi en annexe page 36 par l'algorithme 2. Dans ce deuxième algorithme, les valeurs de α_{min} et α_{max} sont des constantes tel que $1 \geq \alpha_{max} > \alpha_{min} \geq 0$. Ces constantes sont là afin de s'assurer que la valeur de α retournée est bien entre 0 et 1. Mais elles offrent aussi la possibilité de ne pas être en présence d'une fonction de masse catégorique ou d'une ignorance totale lorsqu'elles ont une valeur différente de 0 et 1.

Considérant le fait que nous pouvons avoir un nombre important de sources à combiner nous n'utilisons pas l'opérateur de Yager. Car avec cet opérateur, un nombre important de sources peut engendrer davantage de conflit et il peut en résulter une masse trop importante sur l'ignorance. Or ici, afin d'obtenir la réflexion de du contributeur c nous combinons l'ensemble de ses réponses aux questions q. Pour un faible nombre de questions (entre 1 et 5) il serait envisageable d'utiliser l'opérateur de Yager. Mais dans l'objectif de notre modélisation nous considérons que celle-ci peut s'appliquer sur l'ensemble des plateformes de *crowdsourcing*, sans distinction du nombre de questions posées aussi nous considérons la moyenne des masses (20) sur l'ensemble des questions : $m_c^{\Omega_3}$.

5.2.3 Expertise

Nous pouvons finalement aboutir dans cette section à la modélisation de l'expertise d'un contributeur. Plutôt qu'établir un degré d'expertise, nous attribuons une classe d'expertise aux contributeurs comme le font Abassi et Boukhris [24]. Pour ce faire, nous considérons l'échelle de profil suivante inspirée de l'échelle de Koulougli *et al.* [1] :

- *Spammer* : contributeur sans qualification pour la tâche, ses réponses sont instinctives. Le *spammer* n'est pas consciencieux dans la réalisation de la tâche, motivé uniquement par la rémunération de celle-ci il répond aléatoirement et rapidement aux questions.
- Non-Compétent : contributeur sans qualification pour la tâche mais réalisant celle-ci avec réflexion. Le contributeur Non-Compétent est volontaire dans la réalisation de la tâche ce pourquoi il prend le temps de la réflexion dans ses réponses. Néanmoins dû à un manque important de qualification ses réponses ne sont pas exploitables.
- Compétent : contributeur qualifié pour la tâche prenant le temps de la réflexion lors de la réalisation de celle-ci.
- Expert : contributeur qualifié dont les réponses sont instinctives. L'expert à une excellente connaissance du sujet et nécessite donc un temps de réflexion moindre comparé aux autres contributeurs.

Afin d'associer un profile à un contributeur nous mesurons l'expertise de celui-ci. Nous définissons le cadre de discernement de l'expertise d'un contributeur comme le produit des cadres de discernement de la Qualification et de la Réflexion : $\Omega_4 = \Omega_2 \times \Omega_3$. Afin d'établir la masse $m_c^{\Omega_2 \times \Omega_3}$ nous com-

mençons par définir les masses $m_c^{\Omega_2 \uparrow \Omega_2 \times \Omega_3}$ et $m_c^{\Omega_3 \uparrow \Omega_2 \times \Omega_3}$ de manière analogue à l'équation (7). Nous appliquons ensuite l'opérateur conjonctif de Yager (6) à ces deux masses afin d'obtenir $m_c^{\Omega_2 \times \Omega_3}$.

Finalement nous appliquons une probabilité pignistique à la masse $m_c^{\Omega_2 \times \Omega_3}$ afin d'établir le profil d'un contributeur. Par exemple, si le maximum de probabilité pignistique est sur l'élément $\{Q, R\}$ le contributeur est classifié comme Compétent.

La section suivante explicite les expérimentations que nous avons réalisées afin de valider notre modèle.

6 Validation de la modélisation

Les tests de la modélisation proposée sont faits sur des données issues d'une campagne de *crowdsourcing*. Cette campagne est d'abord présentée ici après quoi, la validation de notre modèle est abordée.

6.1 Présentation de la campagne de *crowdsourcing*

La campagne dont nous exploitons les résultats a été réalisée dans le cadre d'une étude en collaboration avec la société Orange sur la plateforme de *crowdsourcing* d'activités routinières Foule Factory. Cette plateforme se considère comme une ambassadrice du *crowdsourcing* en France.

La tâche à réaliser consistait en l'écoute d'enregistrements sonores et la notation de ceux-ci suivant l'échelle : Excellent, Bon, Correct, Pauvre, Mauvais. Cette campagne est constituée de 4 HITs (Human Intelligence Tasks) comprenant 12 questions chacun parmi lesquelles figurent 5 MNRUs (Modulated Noise Reference Units). Les MNRUs sont des enregistrements dont la qualité sonore est modulée par du bruit. Il s'agit ici de données que l'on peut considérer comme un corpus de référence dont on connaît la qualité relative, échelonnée de Mauvais (1er enregistrement) à Excellent (5ème enregistrement). La différence de qualité sonore entre les MNRUs est beaucoup plus importante que celle qui existe entre les 7 autres enregistrements, nommées par la suite données de test. La qualité des données de tests n'est pas connue et plus difficile à évaluer. Une foule de 93 contributeurs devait écouter chacun des enregistrements dans un ordre aléatoire et spécifier leur qualité. Leurs réponses pouvaient être imprécises, avec la possibilité de donner deux niveaux consécutifs de qualité, et étaient associées à un degré de certitude indiquant la confiance sur le résultat fourni : Très sûr, Plutôt sûr, Moyennement sûr, Peu sûr, Pas sûr.

La figure 10 page 38 en annexe illustre l'interface avec laquelle le contributeur interagît lors de la réalisation de la tâche. Pour chaque enregistrement l'utilisateur a la possibilité de réécouter celui-ci. Afin de s'assurer de l'attention du contributeur dans la réalisation de sa tâche, chaque HIT inclut en plus des 12 enregistrements, une question d'attention. Avant de réaliser sa tâche le contributeur devait renseigner certaines données à des fins statistiques (qui ne seront pas abordées ici) et réaliser un test de calibrage de son matériel audio ainsi qu'une session d'entraînement à la tâche.

Une analyse préliminaire des données est réalisée avant d'établir notre modélisation et l'expérimentation de celle-ci. Nous nous sommes dans un premier temps intéressés à la confiance de l'utilisateur. La figure 2 représente, pour l'ensemble des degrés de confiance que peut spécifier l'utilisateur, le nombre total de fois où cette confiance à été spécifié sur l'ensemble des réponses aux questions. On

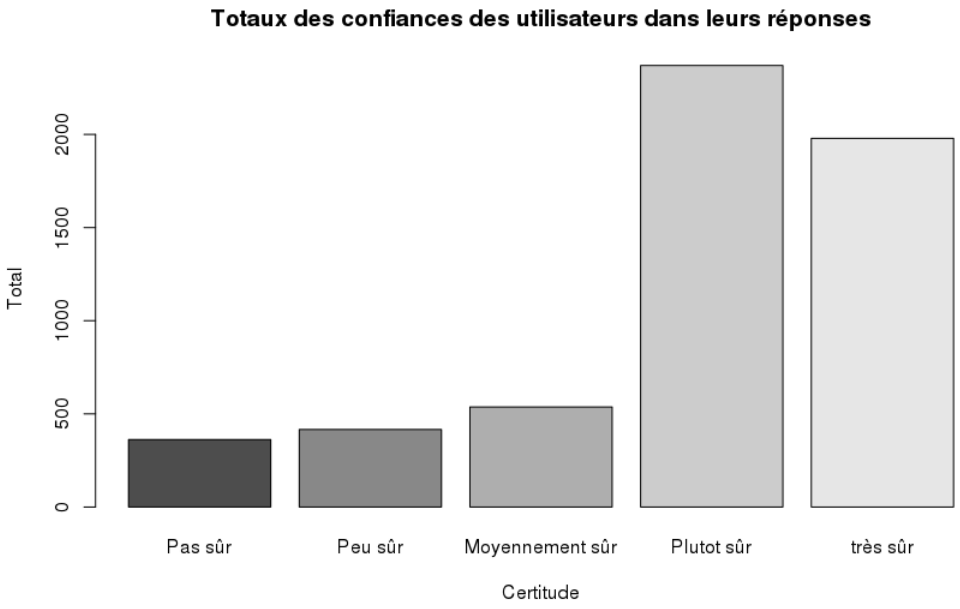


FIGURE 2 – Totaux des confiances des utilisateurs sur l'ensemble des réponses

constate graphiquement que les utilisateurs sont majoritairement "Plutôt sûr" voire "Très sûr" de leurs réponses. Ces degrés de confiance sont gradués tels que "Très sûr" correspond à une confiance de 5 et "Pas sûr" à une confiance de 1 afin de mesurer la confiance moyenne des utilisateurs pour chaque question de chaque hit (figure 11 en Annexe). Enfin, la moyenne sur les graphiques de la figure 11 en annexe est réalisée afin d'obtenir la figure 3. Nous constatons sur cette figure que la confiance moyenne du contributeur lors de la réalisation de sa tâche se situe entre 4 et 5, soit entre "Plutôt sûr" et "Très sûr". L'enregistrement $n°1$ est celui pour lequel la confiance des contributeurs en leur réponse est la plus haute. Cette confiance plus importante s'explique par le fait que cet enregistrement est celui du premier MNRU. Or pour cet MNRU la qualité sonore est plus dégradée que pour les autres et donc plus facilement identifiable.

Un intérêt de cette campagne était que l'utilisateur avait la possibilité d'être imprécis dans ses réponses. Nous nous sommes alors intéressés au pourcentage d'utilisateurs qui ont été imprécis pour chaque question de chaque hit (figure 12 en annexe). Puis la moyenne de ce pourcentage est réalisée afin d'obtenir la figure 4. On constate sur ce graphique qu'excepté pour l'enregistrement $n°1$, entre 20 et 30% des contributeurs ont utilisé la possibilité d'être imprécis. Ce pourcentage bien que faible apparaît comme positif, car les contributeurs ne sont pas accoutumés à être imprécis dans leurs réponses. En effet, dans une campagne passée similaire à celle-ci étudiée par l'équipe DRUID, les contributeurs n'avaient pas bien assimilé la possibilité d'être imprécis et n'utilisaient pas pour la grande majorité cette possibilité. De plus les contributeurs ont mal assimilés dans cette précédente campagne les notions d'incertitude et d'imprécision, aussi lorsqu'ils étaient imprécis ceux-ci avaient une confiance moindre en leurs résultats. Ce qui est paradoxale car en étant moins précis il y a davantage de chance que la bonne réponse se trouve parmi celles renseignées par le contributeur. Or dans la campagne réalisée ici pour tester nos résultats davantage d'utilisateurs ont utilisé la

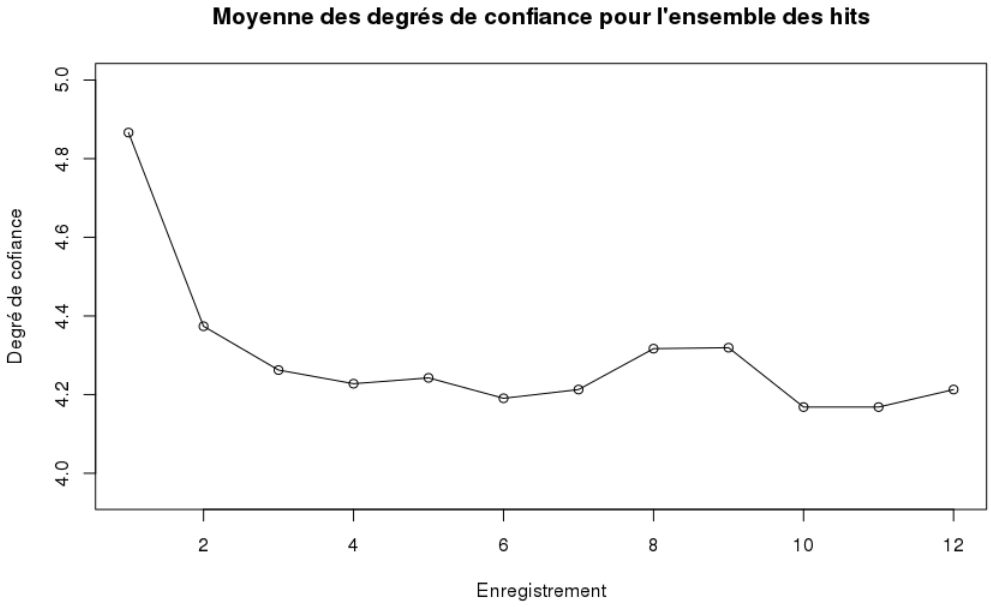


FIGURE 3 – Moyenne des degrés de confiance pour l'ensemble des hits

possibilité d'être imprécis, et n'ont pas fait l'amalgame entre incertitude et imprécision puisqu'ils restent confiant en leurs réponses. Pour l'enregistrement $n°1$ le pourcentage d'utilisateur imprécis est très faible (moins de 10%), ce qui s'explique par le fait qu'il s'agisse du MNRU 1 et concorde avec l'importante confiance des utilisateurs en leur réponse (figure 3).

La confiance avérée des utilisateurs dans leurs réponses est positives pour l'exploitation des données. De plus, dans la campagne précédente, les contributeurs avaient également la possibilité d'être imprécis mais n'utilisaient pas cette imprécision la majorité du temps, ou lorsqu'ils l'utilisaient ils se considéraient comme moins confiant dans leurs réponses. Or la possibilité d'être imprécis devrait renforcer la confiance de l'utilisateur en sa réponse, car il est préférable d'avoir une réponse imprécise et certaine plutôt qu'une réponse précise mais incertaine. Dans cette campagne ce lien entre imprécision et certitude est mieux assimilé par l'utilisateur. La section suivante renseigne des informations sur la bibliothèque et les ressources utilisées pour l'implantation de la modélisation.

6.2 Bibliothèques et ressources

L'implantation est faite en langage R sous l'environnement de développement Rstudio. La bibliothèque *ibelief* est essentielle pour l'implantation, elle permet l'utilisation de fonctions dans le cadre de la théorie des fonctions de croyance telles que : *DST* qui applique l'opérateur conjonctif de notre choix à un ensemble de fonctions de masse, et *mtobetp* qui transforme une fonction de masse renseignée en une probabilité pignistique.

Pour l'équation 16 il est nécessaire de calculer une distance de Jousselme, pour ce faire, un algorithme pour cette distance, réalisé par le passé dans le cadre de tests au sein de l'équipe est

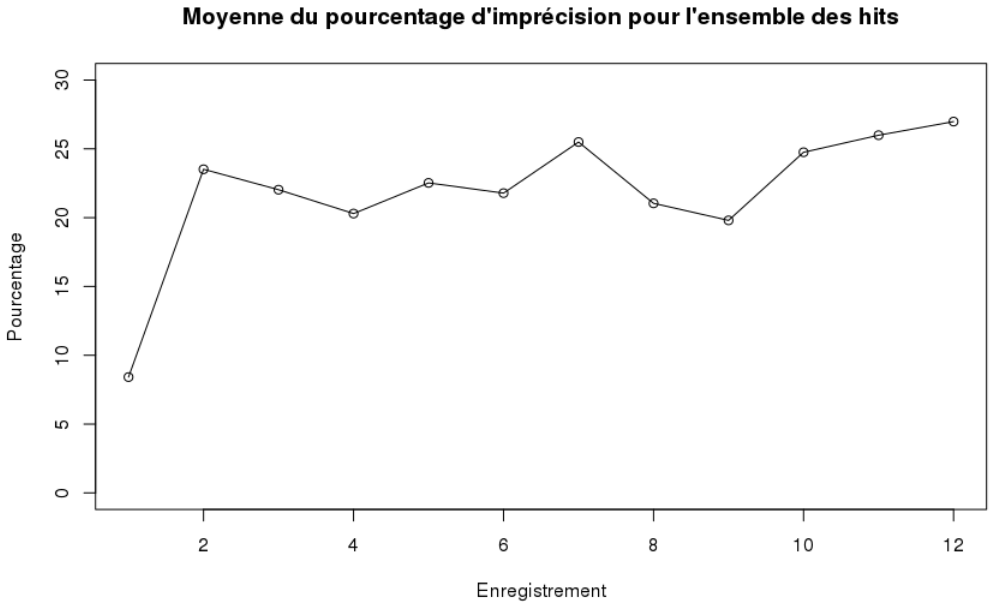


FIGURE 4 – Moyenne des pourcentages d'imprécision des utilisateurs pour l'ensemble des hits

réutilisé. De plus, les opérateurs d'extension vide et de marginalisation ne sont pas définis dans la bibliothèque *ibelief*, ils ont donc été implantés au cours du stage.

Après cette présentation des ressources utilisées pour l'implantation de la modélisation, nous allons maintenant aborder les résultats de celle-ci.

6.3 Résultats de la modélisation

Le corpus de référence (les MNRUs) a été considéré pour la validation de la modélisation uniquement. En effet, nous utilisons ce corpus pour valider nos résultats sur la modélisation de la réponse du contributeur ainsi que pour l'estimation de la connaissance de celui-ci.

Nous commençons ici par calculer la fonction $m_{c\_q}^{\Omega_1}$, pour ce faire nous associons les réponses sur la notation de la qualité d'un enregistrement au cadre de discernement Ω_1, nous avons :

$$\Omega_1 = \{Excellent, Bon, Correct, Pauvre, Mauvais\}$$

La masse α_{cq} associée à cette fonction dépend de la certitude de l'utilisateur, le tableau 3 reprend les valeurs numériques que nous associons aux degrés d'incertitude possibles de l'utilisateur dans le cadre de notre étude. Pour ces valeurs de α_{cq} nous avons choisi de ne pas associer une valeur de 1 à α_{N5} à une réponse indiquée comme "Très Sûr" (réciproquement $\alpha_{N1} \neq 0$ pour une réponse "Pas Sûr"). Car même si le contributeur accorde une confiance absolue dans sa réponse, nous préférons maintenir une incertitude en l'absence de connaissance relative à l'expertise du contributeur. *Via* la mesure de la confiance du contributeur en ses réponses nous pouvons maintenant estimer sa connaissance.

Réponse	α_{cq}
Très Sûr :	$\alpha_{N5} = 0.99$
Plutôt Sûr :	$\alpha_{N4} = 0.75$
Moyennement Sûr :	$\alpha_{N3} = 0.5$
Peu Sûr :	$\alpha_{N2} = 0.25$
Pas Sûr :	$\alpha_{N1} = 0.01$

TABLE 3 – Incertitudes et valeurs associées

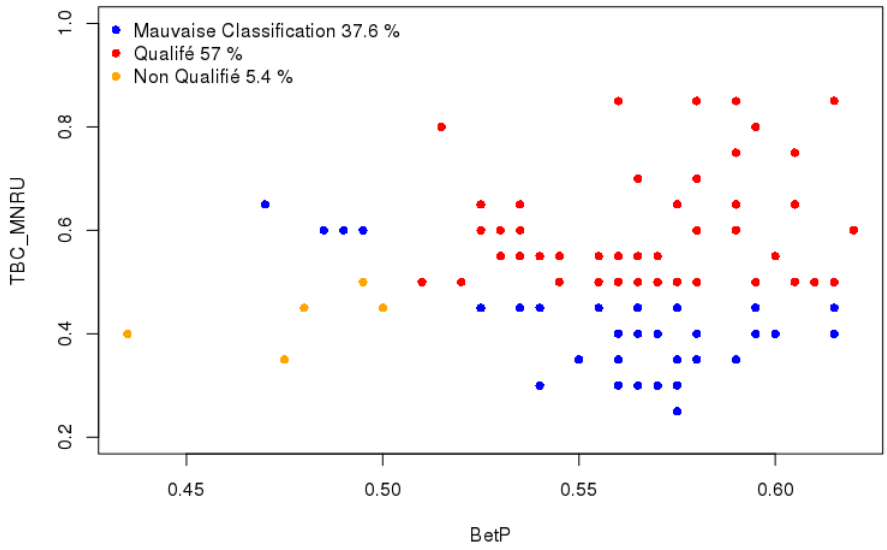


FIGURE 5 – BetP sur les MNRUs

Nous ne connaissons pas le comportement et la connaissance réels de l'utilisateur. Afin d'avoir une estimation théorique de cette connaissance nous utilisons les MNRUs et étudions la convergence des différents résultats obtenus. Pour établir cette estimation théorique, une matrice de confusion est réalisée afin de mesurer les écarts entre les valeurs attendues pour les MNRUs et les réponses attribuées par les contributeurs à l'écoute des MNRUs. Cette matrice nous permet de calculer le taux de bonne classification du contributeur sur les 5 MNRUs (TBC_{M5}) ce qui s'apparente à un taux de bonne réponse pour le contributeur. Nous considérons un seuil $\sigma = 0.5$ au-dessus duquel un contributeur est considéré comme "Qualifié" théoriquement.

Nous appliquons une probabilité pignistique donnée par l'équation (10) à la fonction de masse $m_c^{\Omega_2}$. Nous considérons plus particulièrement la probabilité pignistique sur l'élément $Q \in \Omega_2$ notée BetP sur la figure 5 en ne considérant que les réponses du contributeur aux MNRUs et la figure 6 pour les réponses aux données de test. Pour classifier expérimentalement le contributeur comme "Qualifié" nous considérons le maximum de probabilité pignistique sur Ω_2, ce qui se traduit part $BetP \geq 0.5$. Dans le cas où $BetP < 0.5$ le contributeur est classifié comme "Non-Qualifié".

Considérant les figures 5 et 6 nous constatons un nombre plus important de contributeurs classifiés comme "Qualifié" que "Non-Qualifié" par notre modélisation. Comme nous n'utilisons pas de données d'or pour établir la qualification des contributeurs, notre taux de bonne classification de 62.4% pour la figure 5 et celui de 57% pour la figure 6 nous apparaît positif. En effet, comme la

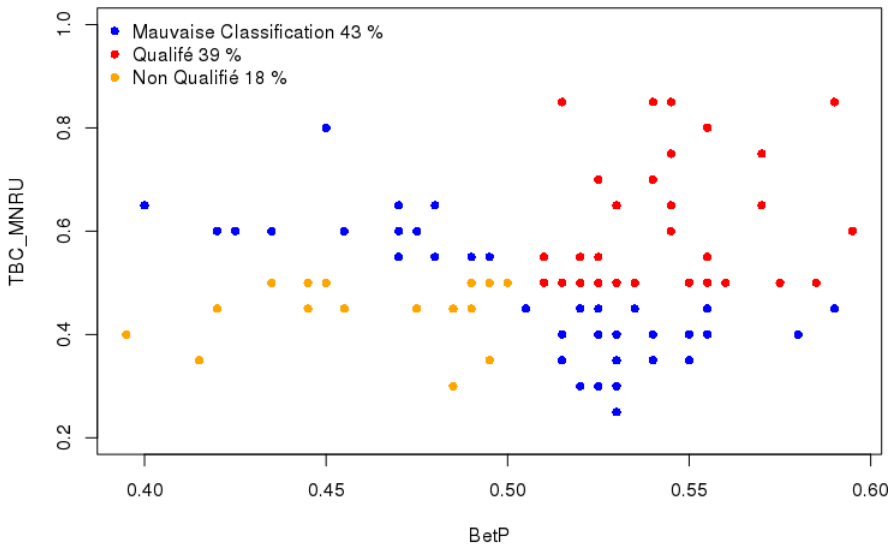


FIGURE 6 – BetP sur les données de test

tâche dépend de l'appréciation du contributeur et que les fluctuations dans la dégradation sonore des données n'est pas évidente à percevoir, ces éléments peuvent expliquer la différence entre les résultats attendus pour les données d'or et ceux obtenus. De plus si nous considérons les résultats de classification de la figure 5 comme référence et que nous comparons leur intersection avec ceux de la figure 6, nous avons 85.85 % de bonne classification en commun. Cela signifie que nos résultats restent à un bon niveau de fiabilité en considérant des questions différentes, MNRUs ou données de test.

La mesure de la Qualification du contributeur ($m_c^{\Omega_2}$) proposée reposant sur les degrés DE_c donnés par l'équation (16) et DP_c par l'équation (17) de Ben Rjab *et al.* [11], nous comparons également les résultats de notre modélisation avec ceux obtenus par ce modèle donné par l'équation (18). Pour ce faire, le degré globale d'expertise du contributeur DG_c est considéré avec la valeur de β la plus appropriée d'après les travaux de Ben Rjab *et al.* : $\beta_c = 0.5$. En mesurant ce degré global DG_c pour l'ensemble des MNRUs nous obtenons la figure 7. Nous pouvons constater avec le même seuil $\sigma = 0.5$ que la majorité des contributeurs (88.2%) est considérée comme "Non-Expert" par ce modèle. Nous obtenons pour le degré DG_c un taux de bonne classification pour l'estimation de la qualification du contributeur de 60.6% pour les MNRUs et de 58.0% pour les données de tests.

Le taux de bonne classification pour la qualification du contributeur $m_c^{\Omega_2}$ est plus important pour les MNRUs que pour les données de test. Cette différence est due à la différence d'échelle de dégradation sonore plus importante pour les MNRUs que pour ces autres données. Le même raisonnement s'applique au degré DG_c, car ces mesures reposent sur les degrés DE_c et DP_c, ce qui explique la diminution de leur taux entre les MNRUs et les données de tests. Les taux de bonne classification de ces deux mesures sont proches. Malgré la difficulté de proposer une modélisation pertinente de la connaissance des contributeurs en l'absence d'un corpus de référence, notre approche apporte des résultats intéressants à la vue des expérimentations menées.

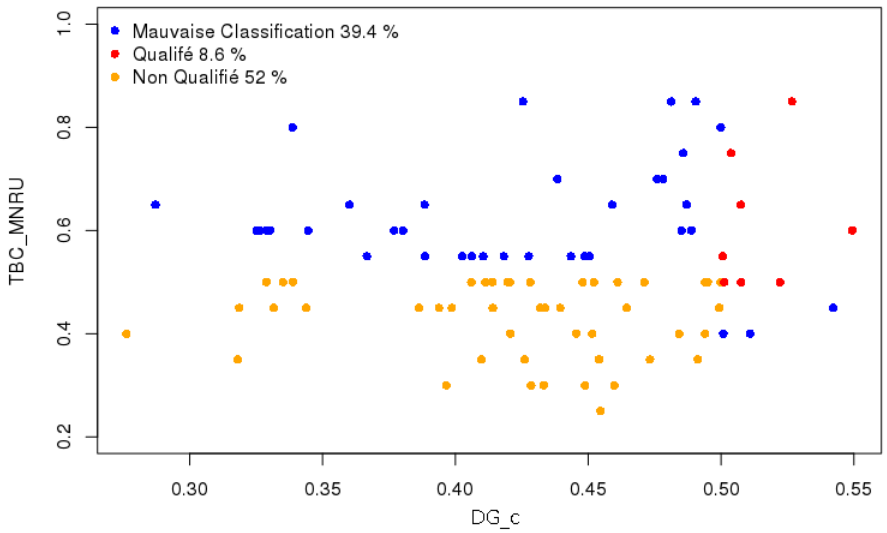


FIGURE 7 – DG_c sur les MNRUs.

De manière analogue à la connaissance nous ne connaissons pas le comportement véritable des contributeurs dans la réalisation de la tâche. Nous réalisons une estimation théorique du comportement considérant le temps de réponse du contributeur et un temps de réponse théorique attendue aux questions. Pour obtenir ce temps de réponse théorique (T_{rep_th}), nous sommons l'ensemble des temps des enregistrements (T_{enr}) des MNRUs. Nous considérons arbitrairement que l'utilisateur prend 10 secondes pour réfléchir à sa réponse, à la certitude qu'il a en celle-ci et pour renseigner ses informations. En ajoutant ce temps de réponse de référence pour les 5 MNRUs des 4 HITs à la somme des temps des enregistrements nous obtenons T_{rep_th}. Pour chaque contributeur nous comparons la somme de ses temps de réponse sur les MNRUs (T_c) à T_{rep_th}. Si $T_c > T_{rep\_th}$ les réponses de l'utilisateur sont considérées comme théoriquement "Réfléchie" (R), à l'inverse, si $T_c \leq T_{rep\_th}$ les réponses sont considérées comme "Instinctives" (NR). De plus, dans l'algorithme 2 nous avons : $\alpha_{min} = 0.01$ et $\alpha_{max} = 0.99$ afin de ne pas avoir de fonctions de masse catégoriques ni d'ignorance totale, ainsi on préserve une faible incertitude sur le comportement du contributeur.

Une probabilité pignistique est appliquée à la fonction de masse $m_c^{\Omega_3}$ afin de classifier les contributeurs réfléchis (R) et les instinctifs (NR). En comparant notre classification sur le comportement du contributeur pour les MNRUs à la classification de référence établie nous avons un taux de bonne classification de 51.6%. Pour les données de test ce taux est plus faible : 41.9%. Ce qui apparaît paradoxal ici c'est que 64 contributeurs sont classifiés comme Réfléchis par notre modélisation pour les MNRUs et seulement 57 pour les données de test.

Or il est plus facile de les distinguer les MNRUs en terme de qualité d'écoute car leur dégradation sonore est plus prononcée que pour les données de test. Les contributeurs devraient donc théoriquement prendre d'avantage de temps de réflexion pour les données de tests ce qui n'est pas le cas ici. De plus cette diminution du nombre de contributeurs "Réféchis" s'accompagne de la diminution du taux de bonne classification, ce qui confirme l'hypothèse que d'avantage de contributeur devraient être estimés réfléchis que ce qui a été établie par la classification. Ceci s'explique par le fait que la majo-

MNRUs	Experts	Compétents	*Spammers*	Tous	MV
1	1	1	1	1	1
2	2	2	3	2	1
3	3	{1,2}	2	3	4
4	4	{1,2,4}	1	4	4
5	4	4	2	4	4

TABLE 4 – Résultats sur les notes pour la modélisation proposée et pour la méthode par vote (MV)

rités des contributeurs (les compétents et non-compétents) doit prendre un temps de réflexion pour répondre. Les contributeurs répondant de manière instinctive que se soit parce qu'ils sont fortement qualifiés et efficaces dans la réalisation de la tâche (les experts) ou qu'ils répondent aléatoirement (les *spammers*) sont a priori moins nombreux.

Le comportement du contributeur relève de la motivation de celui-ci dans la réalisation de la tâche. Or il n'existe à ce jour aucune méthode théorique permettant de caractériser le comportement du contributeur. Aussi, bien que le taux de bonne classification pour notre modélisation du comportement est mitigé, celui-ci reste positif en l'absence de méthode existante.

Après avoir calculé les fonctions de masse sur les cadres Ω_2 et Ω_3 nous pouvons finalement aboutir à la mesure de l'expertise du contributeur sur le cadre de discernement $\Omega_2 \times \Omega_3$. Encore une fois nous n'avons pas de connaissance réelle de l'expertise des contributeurs. Afin d'analyser les résultats sur l'expertise, la moyenne des réponses des contributeurs aux MNRUs est réalisée en regroupant les réponses par classe de contributeurs afin d'obtenir les masses m_{q_exp} (ou exp est la classe d'expertise du contributeur). Par exemple nous appliquons l'opérateur de moyenne à l'ensemble des contributeurs caractérisés comme compétents par notre modélisation pour les 5 MNRUs et obtenons la masse m_{q_Qual} pour chaque MNRU q. Ensuite nous appliquons une probabilité pignistique à ces masses m_{q_exp} afin d'obtenir les notes contenues dans le tableau 4. Nous comparons les résultats sur les notes obtenues par le modèle proposé avec celles obtenues en appliquant la méthode par vote majoritaire. Comme aucun utilisateur n'est classifié comme "Non-Compétent" par la méthode de caractérisation des contributeurs, cette colonne n'apparaît pas dans le tableau 4.

Une première constatation faite sur le tableau 4 est que dans aucun cas, le MNRUs de qualité "Excellent" (celui qui a une note de 5), ne se voit attribuer cette note, que se soit avec l'utilisation de notre modélisation ou avec la méthode par vote majoritaire. On en déduit que par un phénomène culturel le contributeur ne parvient pas à attribuer la meilleure note à un enregistrement, quand bien même cela est le cas. Or Leclercq *et al.*[27] dans leur article traitant de la docimologie [11] abordent différents biais de notation. Dans l'ensemble ces biais sont propres à la notation scolaire, cependant, celui portant sur l'effet de tendance centrale semble propice ici. Cet effet de tendance centrale se traduit par une répartition Gaussienne des notes de l'utilisateur. Ici nous avons une répartition Gaussienne de l'appréciation de la qualité de l'enregistrement par le contributeur. En effet, la figure 8, qui est la moyenne des graphiques pour chaque HITs (figure 13 en Annexe), a une forme approximative de Gaussienne centrée sur la qualité :"Bon". De plus, les graphiques pour chaque HITs ont également cette forme approximative de Gaussienne. Nous pouvons supposer que si d'avantage de réponses avaient été proposées pour la qualité de l'enregistrement, cette Gaussienne

11. La docimologie est une science étudiant les contrôles de connaissance en milieu scolaire.

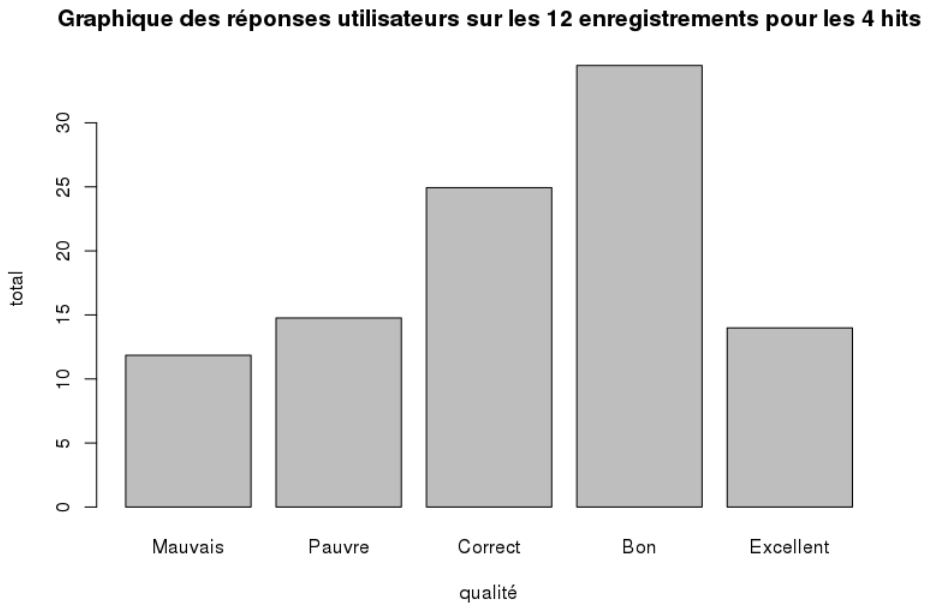


FIGURE 8 – Pourcentage des réponses suivant l'appréciation pour l'ensemble des questions des 4 HITs

serait d'avantage marquée. C'est cet effet de tendance centrale du contributeur qui biaise en partie sa notation et explique pourquoi le MNRU 5 n'est pas qualifié "d'Excellent".

On constate tableau 4 que les meilleurs résultats sur les notes en terme de proximité avec les MNRUs sont ceux offerts par la combinaison des notes pour les contributeurs caractérisés d'Experts. Il est intéressant de remarquer que pour le MNRU 3 (respectivement le 4), pour les utilisateurs Compétents, le maximum de probabilité pignistique est le même pour les notes 1 et 2 (respectivement 1,2 et 4). La notation des contributeurs Compétents, bien que proche de celle attendue par les MNRUs est néanmoins moins pertinente que celle des contributeurs classifiés comme Experts, ce qui est positif car cela permet de valider la modélisation de la caractérisation du contributeur proposée dans la section 5.2.3. A plus forte raison lorsqu'on regarde les notes obtenues par la fusion des réponses des *spammers* qui sont totalement décorélées des réponses attendues.

On remarquera que la probabilité pignistique appliquée à la fusion de l'information pour l'ensemble des contributeurs offre les mêmes résultats sur les notes que pour les contributeurs Experts. Ainsi les réponses ne sont pas impactées négativement par la contribution des *spammers*, notre modélisation sur la réponse du contributeur est donc robuste face à ceux-ci. De plus cette modélisation des réponses offre de bien meilleurs résultats que ceux offerts par la méthode par vote majoritaire.

7 Conclusions et perspectives

Cette section conclue sur l'étude menée ici et propose des axes possibles de recherche dans la continuation des traveaux réalisés dans le cadre du stage.

7.1 Conclusions

Le *crowdsourcing* repose sur l'externalisation de tâches non réalisables par un ordinateur, confiées à une foule de contributeurs. Les tâches étant très diversifiées, on recense différents types de *crowdsourcing*, cette étude s'intéresse au *crowdsourcing* d'activités routinières. Sur les plateformes de ce type, les tâches sont simples et accessibles à tout un chacun, aussi la foule y est la plus diversifiée. Cette grande diversité dans la foule de contributeurs amène à considérer les problématiques suivantes : comment motiver le contributeur dans la réalisation de la tâche, et comment caractériser le contributeur et sa contribution. Nous ne pouvons influer directement sur la motivation de la foule, aussi nous sommes nous focalisés sur la caractérisation du contributeur et de ses réponses. Il est essentiel de caractériser le contributeur pour exploiter au mieux ses réponses, c'est pourquoi, un intérêt particulier est porté à la modélisation des réponses afin d'optimiser leur fusion.

Un état de l'art est réalisé afin d'obtenir une cartographie de l'existant sur les méthodes utilisées dans le cadre du *crowdsourcing* pour la caractérisation des contributeurs et de leurs contributions. Les méthodes utilisant un corpus de référence (données d'or et apprentissage automatique) sont différenciées de celles en faisant abstraction (méthode par vote et approche probabiliste). La méthode traditionnellement utilisée dans les plateformes de *crowdsourcing* est la méthode par vote majoritaire. L'inconvénient majeur de cette méthode est quelle ne tient pas compte de l'incertitude sur la réponse de l'utilisateur, contrairement aux méthodes probabilistes. Aussi, un intérêt particulier a été porté aux méthodes probabilistes utilisant un algorithme EM afin de déterminer la réponse à considérer et l'expertise des contributeurs. Néanmoins elles ne prennent pas en considération l'imprécision sur la réponse du contributeur contrairement à la théorie des fonctions de croyance. C'est pourquoi la modélisation de la caractérisation du contributeur et de sa contribution, exposée dans ce rapport, se fonde sur cette théorie.

La théorie des fonctions de croyance est intéressante dans le cadre de la fusion d'informations incertaines et imprécises. Or dans les plateformes de *crowdsourcing*, considérant les contributeurs comme des sources d'information, les réponses de ces sources sont incertaines et peuvent éventuellement être imprécises. Ben Rjab *et al.* [11] ont étudié sur des données générées, une modélisation faisant abstraction des données d'or et mesurant un degré d'expertise global du contributeur, reposant sur la mesure de degrés d'exactitude et de précision.

Une approche innovente pour la caractérisation des contributeurs et de leurs réponses est proposée dans cette étude. Dans un premier temps, une modélisation sur la réponse du contributeur considérant son incertitude et son imprécision est réalisée. Puis l'expertise du contributeur est modélisée en considérant la connaissance et le comportement de celui-ci. Pour ce faire une échelle d'expertise croissante pour la caractérisation du contributeur est considérée : *spammer*, non-compétent, compétent, expert. Nous exploitons les degrés d'exactitude et de précision de Ben Rjab *et al.* [11] attribué dans cette modélisation à la Connaissance du contributeur. La modélisation du Comportement quant à elle se fonde sur la réflexion du contributeur pour la tâche.

Afin de tester le modèle proposé dans se rapport, une campagne de *crowdsourcing* a été réalisée portant sur la notation d'enregistrements sonores. La possibilité était offerte au contributeur d'être imprécis, et celui-ci devait renseigner la confiance en ses réponses. Une validation des résultats sur la connaissance du contributeur a été faite avec un corpus de référence (les MNRUs). Il est plus difficile de valider le comportement du contributeur en l'absence de connaissance véritable de celui-ci. Néanmoins pour palier à cette difficulté, une estimation théorique du comportement du contributeur est réalisée considérant son temps de réponse. Finalement pour valider les résultats de la classification de l'expertise du contributeur, les réponses sur les MNRUs sont fusionnées suivant la classe du contributeur et les résultats des différentes fusions sont comparées aux MNRUs. De même, pour valider la modélisation sur la réponse du contributeur la fusion de l'ensemble des réponses sans considération de la classe du contributeur est réalisée.

Les expérimentations sur la modélisation de l'expertise concordent en partie à nos attentes et celles sur la modélisation de la note sont très positives. Les erreurs de classification pour l'expertise du contributeur sont attribuées à une estimation incomplète de la Connaissance du contributeur ainsi qu'à un manque d'information sur son Comportement. Afin d'améliorer l'estimation de la Connaissance, la Qualification seule ne suffit pas, aussi il serait intéressant de considérer davantage la tâche et sa difficulté.

7.2 Perspectives

Nous envisageons dans la suite de notre travail afin de parfaire l'estimation de l'expertise du contributeur de lui permettre d'exprimer, non seulement sa confiance sur ses réponses, mais aussi dans quelle mesure la tâche lui paraît intéressante ou complexe. Pour l'intérêt porté à la tâche le cadre de discernement associé serait Ω_I, et les élément de ce cadre appartiendraient à une échelle d'intérêt pouvant aller de "Fortement intéressante" à "Peu intéressante". Cela permettrait de mesurer son intérêt ainsi que son aisance à la mener à bien et apporterait ainsi davantage de renseignements respectivement sur le comportement et la connaissance du contributeur. Considérant la complexité d'une tâche, on peut envisager une échelle de complexité : facile (Nv_1), moyen (Nv_2), difficile (Nv_3). Soit Nv_i le niveau de complexité i d'une tâche, nous pourrions considérer le cadre de discernement $\Omega_C = \{Nv_1, Nv_2, Nv_3\}$. A chaque question cette complexité serait mentionnée moyenne, mais si le contributeur estime être en difficulté face à la tâche il pourrait le renseigner en notifiant que celle-ci est difficile ou dans le cas contraire facile. Ainsi par exemple si la majorité des contributeurs trouveraient la tâche difficile, et que certains contributeurs n'auraient pas eu de difficulté face à celle-ci c'est qu'ils auraient à priori davantage d'expertise. À l'inverse si la majorité des contributeurs estimeraient une tâche facile et que ce ne serait pas le cas pour d'autres, il serait possible de conjecturer que ces contributeurs qui ne trouveraient pas cette tâche facile comme la majorité auraient une qualification moindre pour la tâche. Il serait également intéressant d'observer, en réalisant la moyenne des masses associées à la complexité des questions, si la caractérisation de l'expertise des contributeurs sur la moyenne et celle par rapport à l'ensemble des réponses expliquées précédemment convergent vers une même expertise du contributeur.

Afin d'améliorer l'estimation de la connaissance du contributeur il serait intéressant de savoir s'il est habitué à effectuer des tâches sur les plateformes de *crowdsourcing*. On pourrait considérer l'échelle : habitué "H", moyennement habitué "MH" et novice (non-habitué) "NH". Considérant le cadre de discernement $\Omega_H = \{H, MH, NH\}$ on pourrait par exemple utiliser une fonction de masse

à support simple sur ce cadre. Un contributeur NH serait un contributeur allant sur la plateforme utilisée pour la première fois, un MH aurait déjà utilisé cette plateforme quelques fois (entre 2 et 10 fois) et un H utiliserait celle-ci de manière récurrente (plus de 10 fois). Chose intéressante, l'utilisateur pourrait être considéré comme habitué à une plateforme et moyennement habitué à une autre. Dans ce cas grâce à la théorie des fonctions de croyance il serait possible de le notifier par l'élément $\{H, MH\} \in 2^{\Omega_H}$. Dans le cas où la plateforme de *crowdsourcing* attribue des niveaux à ces contributeurs il serait intéressant d'utiliser ces niveaux comme cadre de discernement. Remarquons que certaines plateformes de *crowdsourcing* attribuent un niveau aux contributeurs suivant leurs qualifications et/ou leurs expériences. Pour exemple, la plateforme FouleFactory propose à ses contributeurs de passer des tests afin de recevoir des "certifications" (sous forme de note) pour effectuer des tâches plus complexes suivant le niveau de certification attendu pour la tâche. Si nous pouvions avoir connaissance de ces niveaux nous pourrions les utiliser comme valeurs théoriques de la connaissance du contributeur.

Pour modéliser le Comportement, nous nous sommes limités à la Réflexion mais il serait intéressant de prendre d'autres critères en compte, tels que l'attention du contributeur, afin de parfaire notre modèle. Par exemple, si la tâche considérée traite de photographie, il serait intéressant de demander au contributeur si parmi les photos précédentes il y en avait une présentant une femme, un bâtiment etc. Ces questions d'attention nous permettrait alors de considérer si le contributeur est attentif "A" ou non "NA" dans la réalisation de la tâche. À cette question d'attention l'utilisateur devrait une fois de plus spécifier sa confiance en sa réponse. Une masse serrait alors associée aux éléments du cadre de discernement $\Omega_A = \{A, NA\}$. Une première masse sur la réponse à la question d'attention serait mesurée de manière analogue à celle sur la réponse du contributeur pour une tâche (section 5.1). Soit le cadre de discernement Θ incluant l'ensemble des réponses possibles à la question d'attention, nous lui associerions pour le contributeur c la fonction de masse à support simple m_c^Θ. Ensuite la distance de joucelme serait mesurée entre la moyenne des masse modélisant la réponse de la foule (excepté pour le contributeur considéré) et la masse associée à la réponse du contributeur : $d_j(m^\Theta_{E_c|c}, m_c^\Theta)$. Il serait alors possible de considérer :

$$\begin{cases} m_c^{\Omega_A}(NA) = \alpha * d_j(m^\Theta_{E_c|c}, m_c^\Theta) \\ \\ m_c^{\Omega_A}(A) = \alpha * (1 - d_j(m^\Theta_{E_c|c}, m_c^\Theta)) \\ \\ m_c^{\Omega_A}(\Omega_A) = 1 - \alpha \end{cases} \tag{21}$$

Dans l'équation (21) la valeur α est présente pour affaiblir la masse afin de préserver de l'incertitude sur l'attention de l'utilisateur. Par exemple, il est possible que l'utilisateur ait été attentif, mais que la question d'attention ait été difficile pour lui, ce pourquoi il serait bon de garder cette incertitude. Par ailleurs α pourrait même être associé à la complexité de la question d'attention. Par exemple considérant l'échelle de complexité mentionnée précédemment, on pourrait avoir les valeurs suivantes : facile = 0.9, moyen = 0.5 et difficile = 0.1. Ainsi de manière générale $\alpha = 0.5$ exepté si le contributeur spécifie une complexité particulière de la question.

Considérant les nouveaux cadres de discernements proposés et ceux déjà étudiés dans ce rapport, la figure 14 en annexe offre un schéma des combinaisons qui pourraient être réalisée sur les différents cadres de discernements afin d'aboutir à la caractérisation de l'expertise du contributeur.

Commençons par expliquer ce schéma pour la modélisation de la connaissance du contributeur. La première ligne, obtention des fonctions $m_q^{\Omega_1}$ et $m_c^{\Omega_2}$, est explicitée dans ce rapport dans la section 5. Lorsque les fonctions $m_c^{\Omega_2}$, $m_c^{\Omega_C}$ et $m_c^{\Omega_H}$ sont obtenues, une extension vide serait faite pour chacune de ces fonctions, puis une combinaison par l'opérateur conjonctif de Yager serait réalisée. Ainsi pour Ω_{Y_1} nous aurions : $\Omega_2 \times \Omega_C \times \Omega_H$.

Pour la modélisation du comportement, de manière analogue à la modélisation de la conaissance, lorsque l'ensemble des fonctions de masses : $m_c^{\Omega_A}$, $m_c^{\Omega_3}$ et $m_u^{\Omega_I}$ seraient établies, une extension vide serait réalisée sur ces fonctions avant de réaliser une combinaison conjonctive de Yager. Ainsi le cadre de discernement Ω_{Y_2} serait définis par le produit cartésien : $\Omega_A \times \Omega_3 \times \Omega_I$.

Les fonctions de masses Ω_{Y_1} et Ω_{Y_2} seraient respectivement l'aboutissement de la modélisation de la connaissance et du comportement du contributeur. Ces deux informations conjointes permettraient de caractériser l'expertise du contributeur. Cette expertise serait alors défini sur le produit cartésiens de cadre de discernement : $\Omega_{Y_1} \times \Omega_{Y_2}$. Afin d'établir ce produit cartésien une extension vide serait réalisée sur les cadres Ω_{Y_1} et Ω_{Y_2} afin de pouvoir ensuite les combiner grâce à l'opérateur conjonctif de Yager.

7.3 Bilan personnel

Au cours du stage, les lignes directrices suivantes des RSE de l'entreprise ont été abordées : relations et conditions de travail, environnement, loyauté des pratiques.

Dans le cadre des relations et conditions au travail, une utilisation appropriée des outils de dialogue interne a été faite, de plus, il n'y a jamais eu au cours du stage de mise en condition de travailleur isolé. S'ajoute à cela, une participation active aux différents séminaires auxquels prenait part l'équipe DRUID, ainsi qu'aux réunions d'équipe et à d'autres réunions inter-équipes apportant des perspectives nouvelles pour le travail du stage. Par exemple des réunion avec l'équipe Loki, dont les travaux de recherche se focalisent sur l'Interaction Homme-Machine (IHM), ont été effectuées pour définir une nouvelle campagne de *crowdsourcing*insistant d'avantage sur les notions d'imprécision et d'incertitude sur la réponse du contributeur. La politique de développement durable de l'IRISA ainsi que celle de l'IUT de Lannion où s'est déroulé le stage ont été respecté. Enfin, les principes de loyauté des pratiques ont été appliqués par le respect de la confidentialité au sein de l'entreprise.

Plus personnellement, j'ai pu découvrir grâce à ce stage de nouveaux aspects du monde de la recherche. Il s'agit d'un milieu de connaissance et de partage en constante évolution, d'où le besoin de réaliser un état de l'art lorsqu'on s'intéresse à un sujet. Or, la réalisation de l'état de l'art par l'étude bibliographique fut un exercice que j'ai beaucoup apprécié et m'a permis de parfaire mes connaissances sur le sujet de stage. Les domaines et possibilités aussi bien en terme de tâches que de traitements des données étant assez vastes sur les plateformes de *crowdsourcing*, un flot important d'informations m'est apparu que j'ai appris à trier, organiser et exploiter.

Ce stage m'a permis de mettre à profits mes connaissances techniques acquises au cours de ma double formation ENSSAT et M2SIF. D'un point de vue personnel il a été très enrichissant et m'a conforté dans ma décision de poursuivre en doctorat.

Annexe

Algorithm 1 Fonction g(réel :T_{c_q}, T_{ref_q}, caractère : X)

caractère $reflexion \leftarrow NR$
réel $masse \leftarrow C^{te}$
réel $\alpha_3 \leftarrow alpha(T_{c_q}, T_{ref_q})$
if $T_{c_q} > T_{ref_q}$ **then**
  $reflexion \leftarrow R$
end if
if $X = reflexion$ **then**
  $masse \leftarrow \alpha_3$
else if $X = \Omega_3$ **then**
  $masse \leftarrow 1 - masse - \alpha_3$
end if
return $masse$

Algorithm 2 Fonction $alpha$(réel : T_{c_q},T_{ref_q})

$\alpha \leftarrow (T_{c_q} - T_{ref_q})/T_{ref_q}$
if $\alpha \leq 0$ **then**
  $\alpha \leftarrow \alpha_{min}$
end if
if $\alpha \geq 1$ **then**
  $\alpha \leftarrow \alpha_{max}$
end if
return α

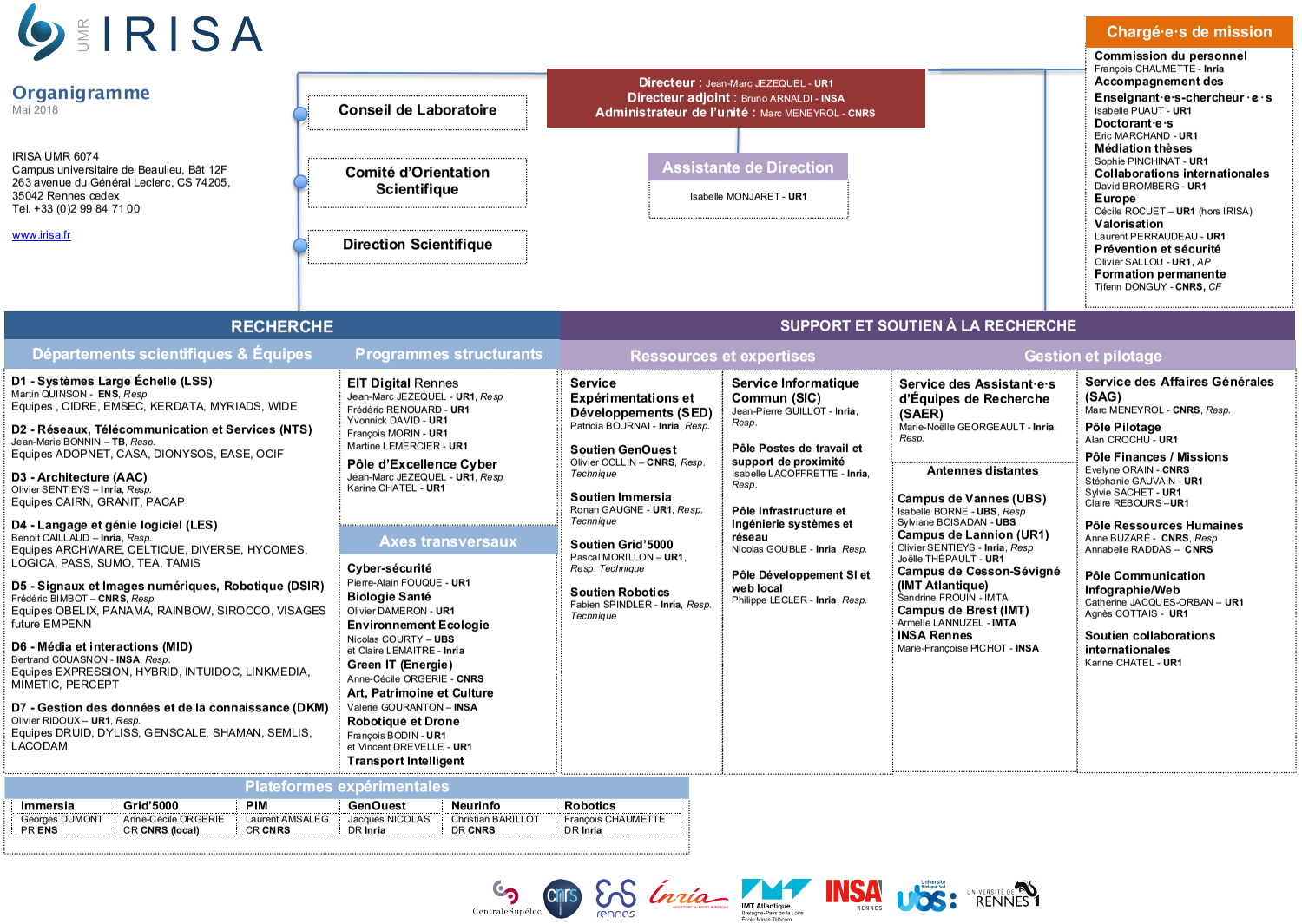


FIGURE 9 – Organigramme de l'IRISA

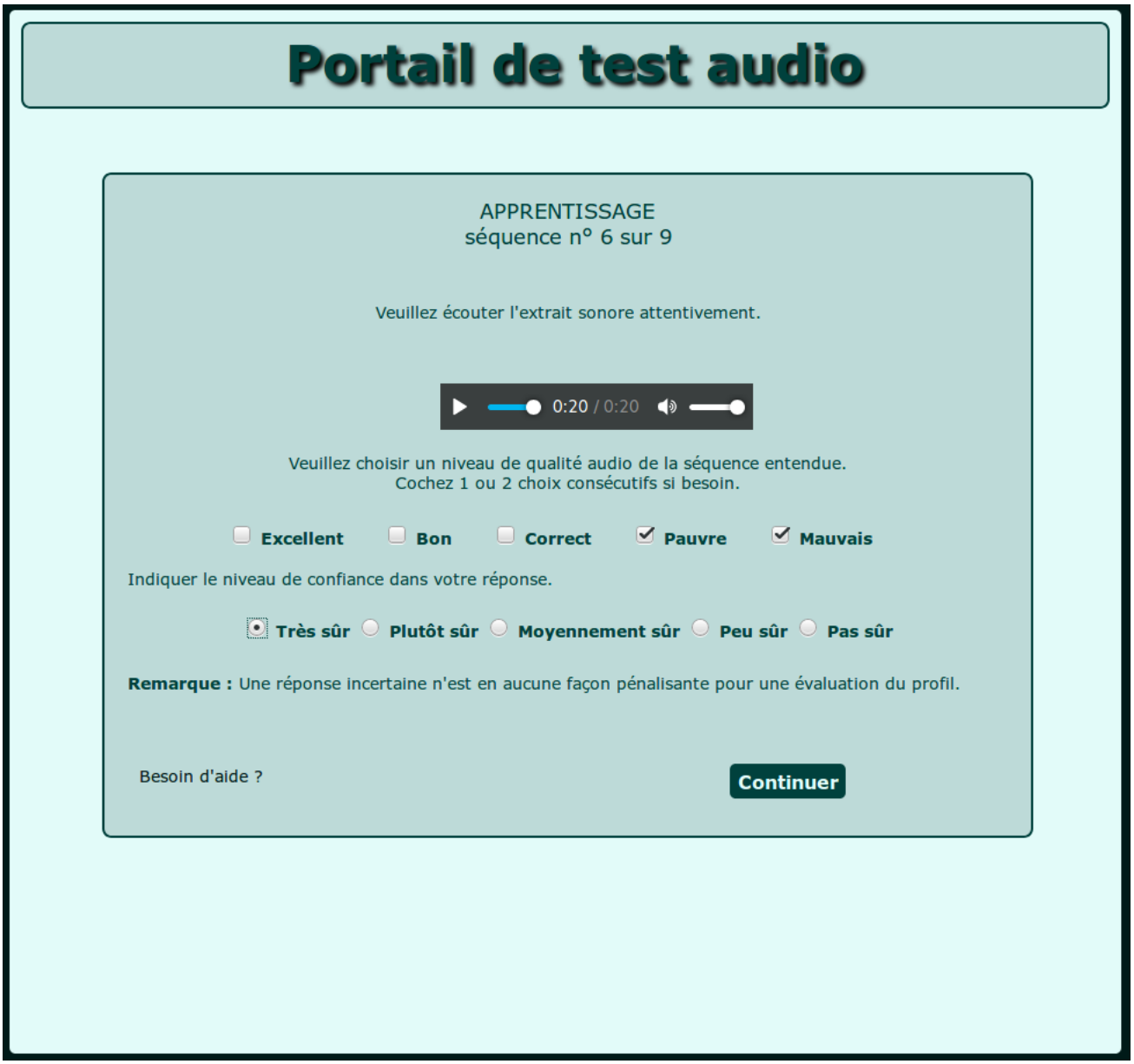


FIGURE 10 – Interface du questionnaire de la campagne de *crowdsourcing*

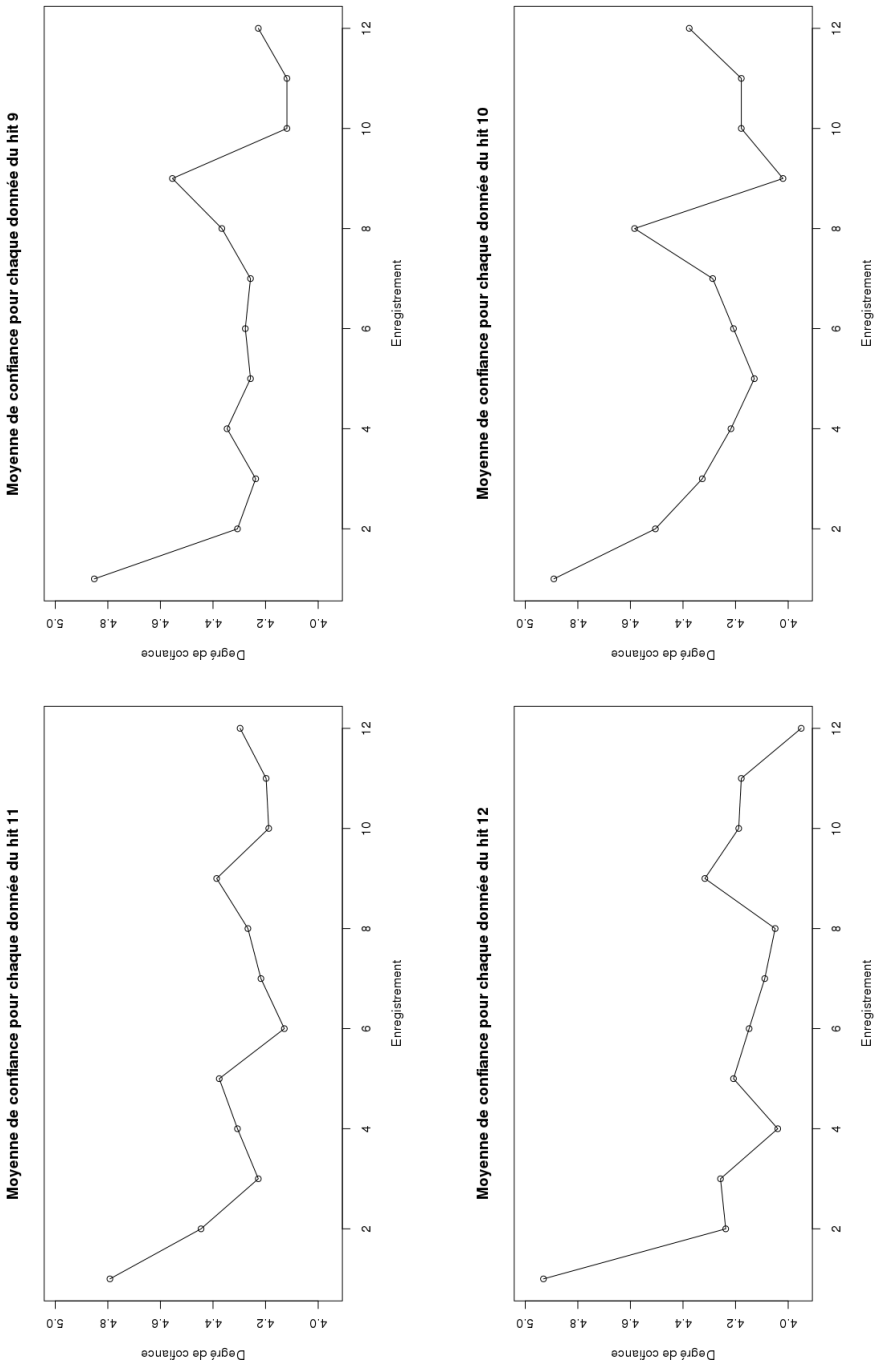


FIGURE 11 – Confiance pour chaque hit

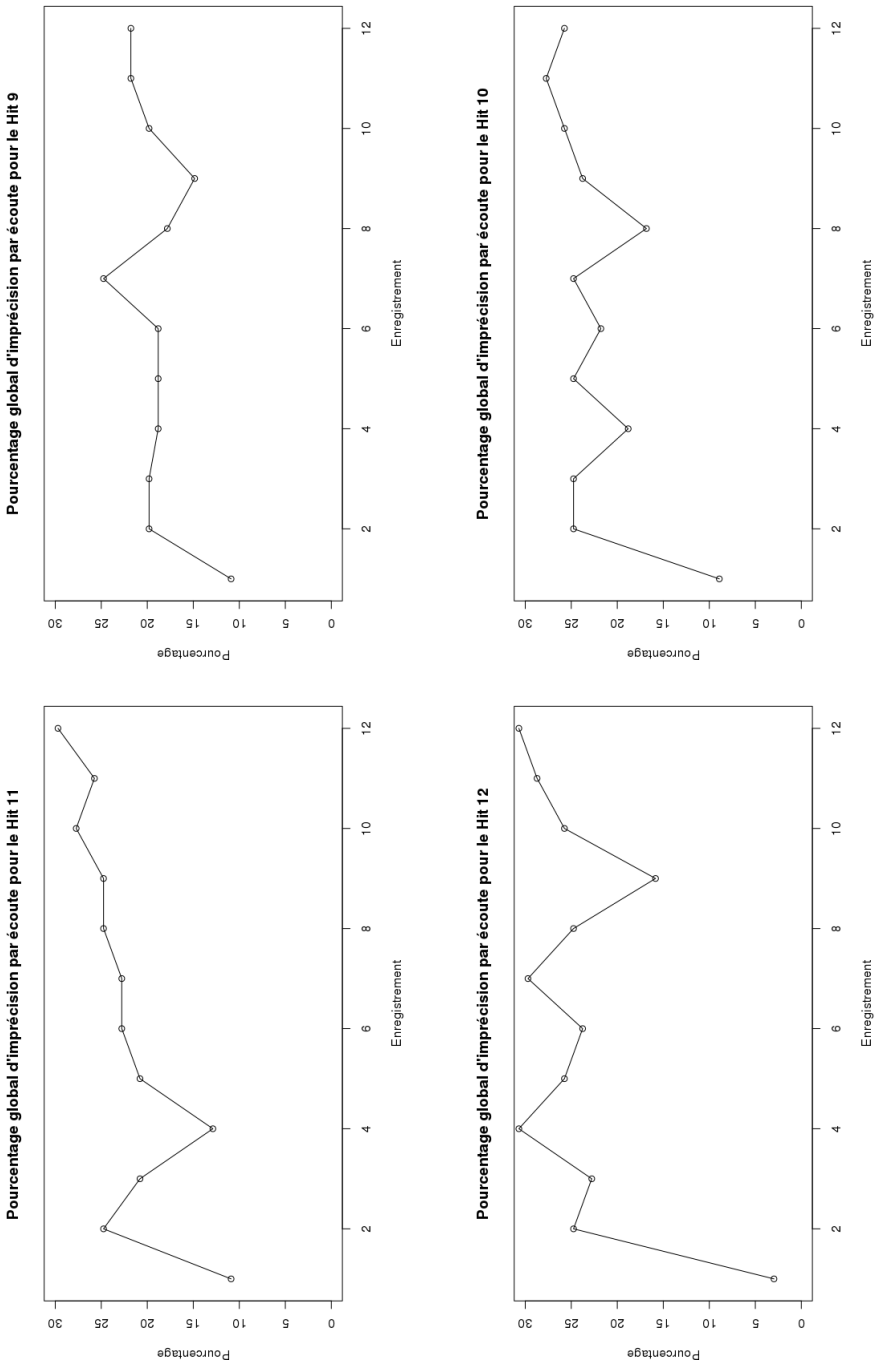


FIGURE 12 – Incertitude utilisateur pour chaque hit

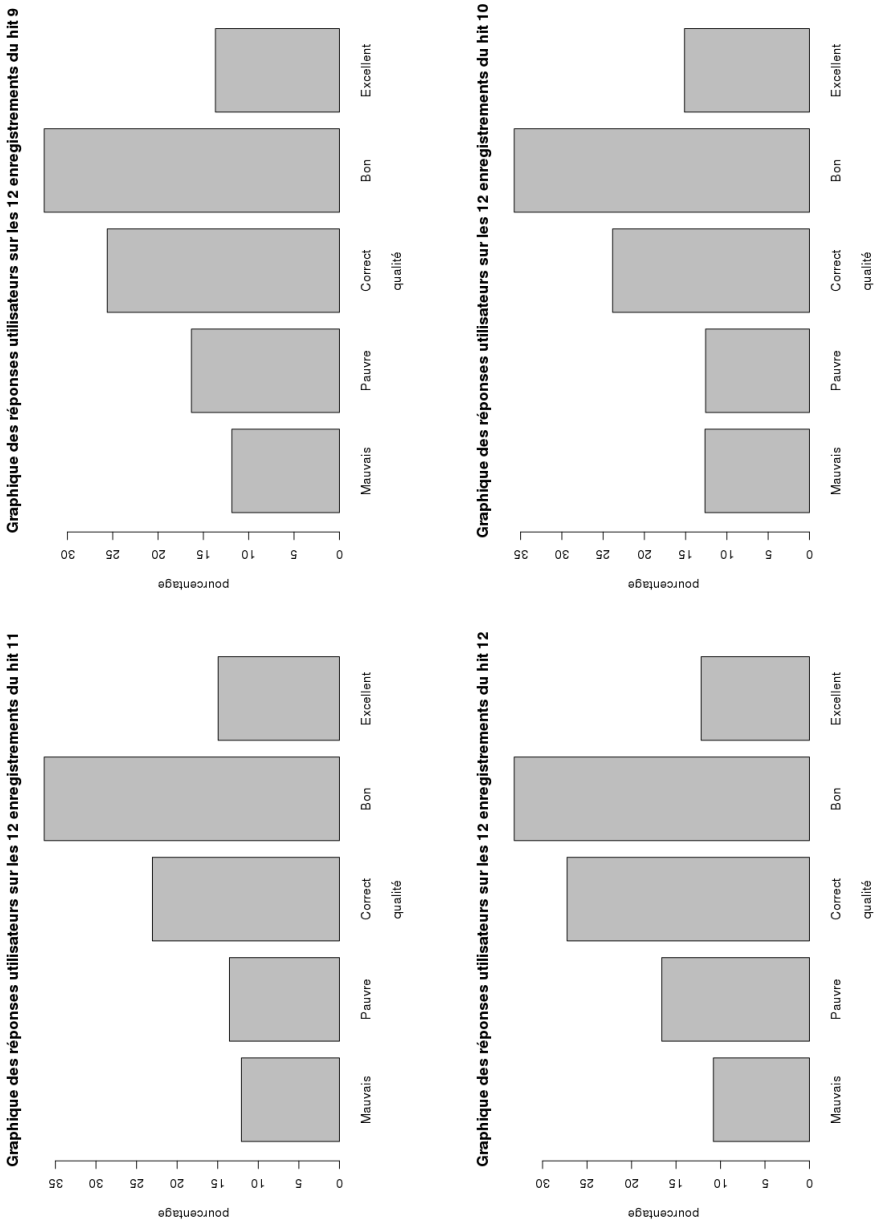


FIGURE 13 – Pourcentage des réponses suivant l'appréciation pour l'ensemble des questions pour chaque hit

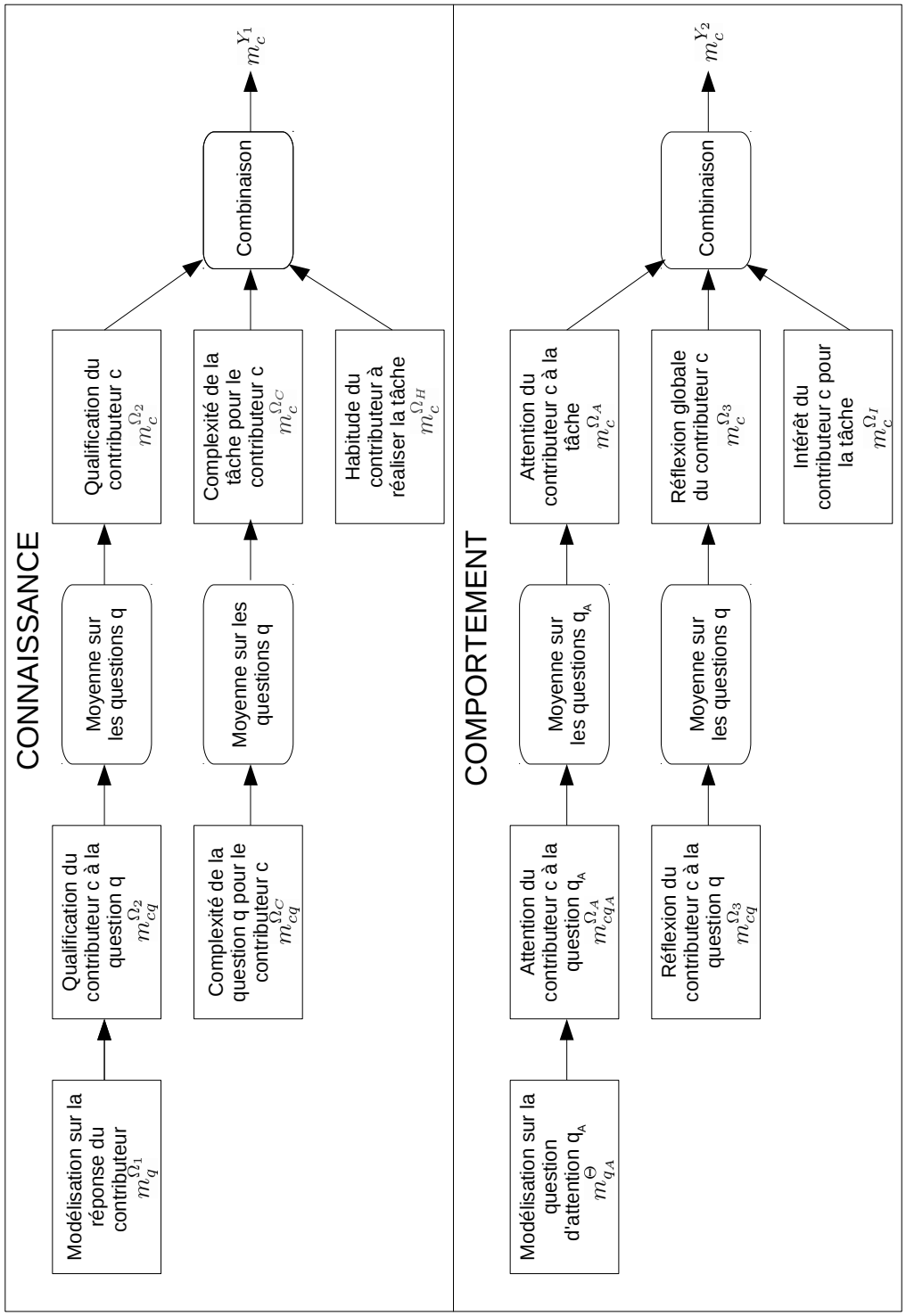


FIGURE 14 – Schéma des perspectives de modélisations et combinaisons possible pour la connaissance et le comportement d'un contribteur

Références

[1] Dalila Kouloughli, Allel Hadjali, and Idir Rassoul. Handling query answering in crowdsourcing systems : A belief function-based approach. In *Fuzzy Information Processing Society (NAFIPS), 2016 Annual Conference of the North American*, pages 1–6. IEEE, 2016.

[2] Vikas C. Raykar ; Shipeng Yu. Annotation models for crowdsourced ordinal data. *Journal of Machine Learning Research*, 2012.

[3] Harry Halpin ; Roi Blanco. Machine-learning for spammer detection in crowd-sourcing. *Human Computation AAAI Technical Report*, 2012.

[4] Vikas C. Raykar ; Shipeng Yu ; Linda H. Zhao ; Gerardo Hermosillo Valadez ; Charles Florin ; Luca Bogoni ; Linda Moy. Learning from crowds. *Journal of Machine Learning Research*, 2010.

[5] Thierry Burger-Helmchen et Julien Pénin. Crowdsourcing : définition, enjeux, typologie. *Management & Avenir*, 41 :254–269, 2011.

[6] Jeff Howe. The rise of crowdsourcing. *Wired Magazine*, 2006.

[7] Alek Felstiner. Working the crowd : Employment and labor law in the crowdsourcing industry. *Berkeley Journal of Employment & Labor Law*, 32 :142–204, 2011.

[8] Aniket Kittur ; Jeffrey V. Nickerson ; Michael S. Bernstein ; Elizabeth M. Gerber ; Aaron Shaw ; John Zimmerman ; Matthew Lease ; and John J. Horton. The future of crowd work. *16th ACM Conference on Computer Supported Coooperative Work (CSCW 2013), Forthcoming*, 2013.

[9] Hosna Ouni ; Arnaud Martin ; Laetitia Gros ; Mouloud Kharoune ; Zoltan Miklos. Une mesure d'expertise pour le crowdsourcing. *Extraction et Gestion des connaissances (EGC)*, 2017.

[10] D. Deutch and T. Milo. Mob data sourcing. *SIGMOD'12*, Tutorial, May 2012.

[11] Amal Ben Rjab ; Mouloud Kharoune ; Zoltan Miklos ; Arnaud Martin. Characterization of experts in crowdsourcing platforms. *Belief Functions : Theory and Applications.*, 9861, 2016.

[12] Jeffrey M. Rzeszotarski and Aniket Kittur. Instrumenting the crowd : using implicit behavioral measures to predict task performance. *UIST '11*, 2011.

[13] Daniel Sabey Weld Peng Dai, Mausam. Decision-theoretic control of crowd-sourced workflows. *Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[14] John Le ; Andy Edmonds ; Vaughn Hester ; Lukas Biewald. Ensuring quality in crowdsourced search relevance evaluation : The effects of training question distribution. *CSE 2010*, 2010.

[15] Gabriella Kazai, Jaap Kamps, and Natasa Milic-Frayling. Worker types and personality traits in crowdsourcing relevance labels. *20th ACM Conference on Information and Knowledge Management, CIKM*, 2011.

[16] Nicolás Della Penna and Mark D. Reid. Crowd & prejudice : An impossibility theorem for crowd labelling without a gold standard. *Collective Intelligence*, 2012.

[17] Alexander Philip Dawid and Allan M Skene. Maximum likelihood estimation of observer error-rates using the em algorithm. *Applied statistics*, pages 20–28, 1979.

[18] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[19] Panagiotis G. Ipeirotis ; Foster Provost ; Jing Wang. Quality management on amazon mechanical turk. *KDD-HCOMP'10*, 2010.

[20] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38 :325–339, 1967.

[21] Ronald R Yager. On the dempster-shafer framework and new combination rules. *Information sciences*, 41(2) :93–137, 1987.

[22] Philippe SMETS. The transferable belief model for expert judgments and reliability problems. *Reliability Engineering & System Safety*, 38, 1992.

[23] Didier Dubois ; Henri Prade and Philippe Smets. Representing partial ignorance. *IEE Transaction on systemes, Man, And Cybernetics - Part A : Systems and humans*, 36, 1996.

[24] Lina Abassi and Imen Boukhris. A worker clustering-based approach of label aggregation under the belief function theory. *Applied Intelligence*, pages 1–10, 2018.

[25] Anne-Laure Jousselme, Dominic Grenier, and Éloi Bossé. A new distance between two bodies of evidence. *Information fusion*, 2(2) :91–101, 2001.

[26] Lewis R Goldberg. The structure of phenotypic personality traits. *American psychologist*, 48(1) :26, 1993.

[27] Dieudonné Leclercq, Julien Nicaise, and Marc Demeuse. Docimologie critique : des difficultés de noter des copies et d'attribuer des notes aux élèves. 2004.