



HAL
open science

Modeling evolutionary responses in crowdsourcing MCQ using belief function theory

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

► To cite this version:

Constance Thierry, Arnaud Martin, Yolande Le Gall, Jean-Christophe Dubois. Modeling evolutionary responses in crowdsourcing MCQ using belief function theory. 27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2023), Sep 2023, Athens, Greece. hal-04214522

HAL Id: hal-04214522

<https://inria.hal.science/hal-04214522>

Submitted on 6 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2023)

Modeling evolutionary responses in crowdsourcing MCQ using belief function theory

Constance Thierry^a, Arnaud Martin^a, Yolande Le Gall^a, Jean-Cristophe Dubois^a

^aUniv. Rennes, CNRS, IRISA, Lannion, France

Abstract

Crowdsourcing is a widespread method for outsourcing tasks to a crowd of contributors. These simple tasks are often formulated by multiple choice questionnaires (MCQs) to which the contributor has to give a precise answer. We hypothesize that offering the contributor to fill in an imperfect answer and to evolve it is more profitable than a single precise answer. In this paper, we propose a model for the evolutionary answers of contributors to MCQs in crowdsourcing platforms. In order to realize this modeling we use the theory of belief functions. The model and experiments conducted on real data from crowdsourcing campaigns are presented in this paper. Our experiments show that modelling evolutionary responses using consonant mass functions improves the quality of the results obtained when aggregating responses compared with majority voting.

© 2023 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the KES International.

Keywords: Belief functions ; questionnaires ; evolutionary responses

1. Introduction

The term crowdsourcing refers to the outsourcing of tasks to a crowd of individuals (contributors) on a dedicated platform, as they cannot be automated and performed by computer, or require human expertise. The diversity of tasks that can be carried out by crowdsourcing leads to a diversity of existing platforms, such as Amazon Mechanical Turk, Crowdspring, or Innocentive. Commonly, the tasks performed on these platforms are simple micro-tasks, achievable without strong expertise, remunerated by a micro-payment. As these tasks are simple, they are often presented in the form of a Multiple Choice Questionnaire (MCQ) where a precise answer is expected from the contributor. Nevertheless, this type of very restrictive answer is penalizing for the contributor in case of indecision. In this respect, Thierry *et al.* [14] has shown that the more imprecise the contributor is, the more certain he is and conversely.

E-mail address: name.surname@irisa.fr

Moreover, it can be interesting to offer the contributor the possibility to evolve his answer to a question. Indeed, Dow et al. [6] allow the contributor to revise his work after an auto-correction or an external feedback. The authors show that both self-correction and external correction allow better quality answers to be obtained. Drapeau *et al.* [7] also evolve the contributor's response through their "MicroTalk" system. In this system, the contributor answers a given question and then provides his reasoning to complement the answer. Finally, the contributor is presented with an argument against the answer and asked to confirm the answer or choose a new one. We call an evolving response a response that will be modified one or several times by the contributor. Factually this evolution can be due to a self-correction as for Dow et al. [6] or a change following a counter-argument as Drapeau *et al.* [7]. In this article, we consider an evolving answer as an answer to which the contributor will bring more precision or in the opposite case, he will enlarge the brought information in order to be more certain in his contribution.

In order to model this evolution in the contributor's response, while preserving the possibility of being imprecise and giving his certainty, we use the theory of belief functions. This theory, introduced by Dempster [3], makes it possible to treat information coming from imperfect sources. Indeed, in crowdsourcing platforms a contributor can be considered as an imperfect source of information due to his human nature.

Our contributions in this paper are as follows:

- modeling of evolutionary responses by consonant mass functions
- comparison of results obtained with the aggregation of evolving and non-evolving responses on real data
- highlighting the impact of the response weighting during aggregation based on its imprecision

The plan of this paper is the following, in section 2 we explain the theory of belief functions necessary for our modeling. In section 3 we present the state of the art for modeling questionnaire responses using the theory belief functions. Our model is then introduced in section 4, followed by a presentation of the experiments conducted in section 5 and finally section 6 concludes this paper.

2. Theory of belief functions

The theory of belief functions, also called Dempster-Shafer theory, because it was introduced by Dempster [3] and formalized by Shafer [11], allows to model the imprecision and uncertainty of imperfect sources of information. We call the set of classes or hypotheses ω_i that are exclusive and exhaustive the frame of discernment $\Omega = \{\omega_1, \dots, \omega_n\}$. The belief functions are defined on:

$$2^\Omega = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_1 \cup \omega_2\}, \dots, \Omega\} \quad (1)$$

The element Ω represents ignorance, and \emptyset symbolizes openness to the world outside the frame of discernment. The belief functions include: credibility, possibility and mass functions, we will focus on the last ones.

The mass functions m^Ω model the elementary degree of belief of the source. They are defined on 2^Ω with values in $[0, 1]$ and respect the normalization condition:

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

The higher the mass $m^\Omega(X)$, the stronger the belief. When $m^\Omega(\emptyset) = 0$ it means that an opening to the world out of the frame of discernment is not possible, one says then to be in closed world. Moreover, for $m^\Omega(\emptyset) > 0$, the function is said to be non-dogmatic. An element $X \in 2^\Omega$ such that $m^\Omega(X) > 0$ is called focal element. We introduced now some specific mass function.

Categorical mass function. When the source is absolutely certain of its answer, all belief is given to $X \in 2^\Omega$. The answer X can be imprecise if it is a meeting of classes belonging to the frame of discernment.

$$m^\Omega(X) = 1, X \in 2^\Omega \quad (3)$$

Simple support mass functions (X^a). The source has an uncertain knowledge because it partially believes in X but not totally since a non-zero mass is present on Ω .

$$\begin{cases} m^\Omega(X) = a \text{ with } X \in 2^\Omega \setminus \Omega, a \in [0, 1] \\ m^\Omega(\Omega) = 1 - a \end{cases} \quad (4)$$

Consonants mass function. The focal elements $X_1, X_2, \dots, X_p \in 2^\Omega$ are nested such that $X_1 \subset X_2 \subset \dots \subset X_p \subseteq \Omega$.

In order to facilitate the combination of mass functions and decreasing conflicts, discounting can be used:

$$\begin{cases} m^{\Omega, \alpha}(X) = \alpha m^\Omega(X), \forall X \in 2^\Omega \setminus \Omega \\ m^{\Omega, \alpha}(\Omega) = 1 - \alpha(1 - m^\Omega(\Omega)) \end{cases} \quad (5)$$

The discounting coefficient $\alpha \in [0, 1]$ models the confidence in the source. The larger α is, the more reliable the source is considered to be. In consequence, if $\alpha = 0$ the source is absolutely unreliable and the totality of the mass is transferred to the ignorance Ω .

Consequently for information fusion, the set of the K sources c all report on the same frame of discernment Ω and the same topic. For all equations defined in this section $X \in 2^\Omega$. It is possible to average mass functions:

$$m_{Avg}(X) = \frac{1}{K} \sum_{c=1}^K m_c^\Omega(X) \quad (6)$$

This simplistic combination of mass functions allows us to remain in a closed world. Many other rules of combination exist in the theory of belief functions, this section lists some conjunctive rules, [10] presents more. Conjunctive combination requires the sources to be reliable, distinct and cognitively independent.

Conjunctive rule. This operator reduces the imprecision on the focal elements and increases the belief on the concordant elements.

$$m_{Conj}^\Omega(X) = \left(\bigcap_{c=1}^K m_c^\Omega \right) (X) = \sum_{Y_1 \cap \dots \cap Y_K = X} \prod_{c=1}^K m_c^\Omega(Y_c) \quad (7)$$

The mass $m_{Conj}^\Omega(\emptyset)$ represents the global conflict of the combination. Other operators allow to remain in a closed world after combination, for example the normalized conjunctive rule of Dempster or the one of [16].

Dempster rule. This operator allows a fair distribution of the conflict $k = m_{Conj}^{\Omega}(\emptyset)$ on the focal elements.

$$m_D^{\Omega}(X) = \frac{1}{1-k} m_{Conj}^{\Omega}(X) \quad (8)$$

Combining the sources of information yields the mass m_{Comb} which is the result of aggregating the mass functions. There are different strategies for making decisions about m_{Comb} in the theory of belief functions among them we present the maximum of pignistic probability $betP$ defined by [12] as:

$$betP(X) = \sum_{Y \in 2^{\Omega}, Y \neq \emptyset} \frac{|X \cap Y|}{|Y|} \frac{m^{\Omega}(Y)}{1 - m^{\Omega}(\emptyset)} \text{ with } X \in 2^{\Omega} \quad (9)$$

The maximum probability is obtained for ω_d such that:

$$betP(\omega_d) = \max_{\omega_i \in \Omega} betP(\omega_i) \quad (10)$$

The section 3 presents the state of the art of the use of the theory of belief functions in a crowdsourcing context.

3. State of the art

Apart from a crowdsourcing context, we can cite the work of Diaz *et al.* [4, 5] for the modeling by the theory of belief functions from answers to MCQs in a university context. The authors point out that MCQs restrict the student to a specific answer even if he is not convinced by the choice. In case of doubt the student has to make a blind choice between several answers that are not necessarily totally wrong. Precise MCQs lack the possibility for the student to express ignorance, imprecision and uncertainty. Diaz *et al.* [5] propose Ev-MCQ, a MCQ allowing students to express a more flexible answer with a mass. Although the context of these articles is the academic world, we find the same problem in the context of crowdsourcing, that is, when faced with a questionnaire that does not allow the contributor to fully express himself, he is restricted in his response, which is disastrous when the contributions are aggregated. Thierry *et al.* [14] show that the more imprecise the contributor is, the more certain he is. The approaches we will now introduce model imprecise and/or uncertain responses in crowdsourcing platforms.

Koulougli *et al.* [9] defined the CASCAD method for modeling and aggregating contributions from crowdsourcing platforms. During the campaign, the participant selects one to several answers and assigns his degree of uncertainty to his contributions, which allows the construction of mass functions. These functions are weakened by a α coefficient related to the expertise of the contributor previously established during a qualification test before the campaign. The responses are then aggregated in cascade by the Dempster combination operator (equation (8)). CASCAD is compared on data generated with the EM algorithm of [2] and the Majority Voting (MV) which is an aggregation method commonly used in crowdsourcing platform. CASCAD has a higher correct response rate than EM and MV when three focal elements are used.

Thierry *et al.* [15] propose MONITOR for estimating the contributor's profile and aggregating the responses. In this model a contributor can choose one to several answers while indicating his certainty. The informed contribution is modelled by a simple support mass function and weakened by the estimated contributor profile. MONITOR is tested on real data from crowdsourcing campaigns and also shows better results compared to majority voting. However, MONITOR does not allow to deal with a possible evolution in the contributor's response.

Unlike the two previous models, the one by Abassi and Boukhris [1] considers precise responses. The accuracy of the contributor is estimated by 3 measures: gold data, MV and log distance. A clustering is performed on these

measures in order to classify the contributors, then a mass is associated to the response of the contributor according to his class. The contributions are then combined. The authors tested their approach on real data.

Although these approaches are interesting, the MCQs used do not allow the contributor to evolve his answer. We consider an evolving answer as a contribution that will be modified for the same question. We propose a model of evolutionary answers by the theory of belief functions in the following section.

4. Modeling evolutionary responses making use of consonant mass functions

When we apply the theory of belief functions in a crowdsourcing framework, a contributor c can be considered as an imperfect source of information due to his human nature. The frame of discernment Ω is the set of possible responses choices ω_i to a question q , and $X \in 2^\Omega$ is the answer set selected by the contributor.

Here we call a dynamic question a question that is repeated to the contributor on the same subject but which may lead him to change his answer. For example, the contributor may be asked to select a set of responses and if the certainty he indicates is low, he may be asked to extend his selection for the same question until a higher level of certainty is reached. When the contributor modifies his answer we make the following assumptions:

- in case the contributor expands his answer X_1 , the new selected answer set X_2 includes the previous answer set i.e. $X_2 \supset X_1$
- in the case where the contributor restricts his answer X_1 , the new response X_2 is a subset of the previous answer's elements i.e. $X_2 \subset X_1$

The contributor may have to change N times his answer selection X but this will always be done as $X_i \subset X_{i+1}$ or $X_{i+1} \subset X_i$. Our definition of dynamic questions and the assumptions made allow us to model the responses collected in this way by consonant mass functions. We denote m_c^Ω the mass function associated with the contributor c for the question q and the frame of discernment Ω .

$$\begin{cases} m^\Omega(X_i) = \delta_i * a_{cqi}, \forall X_i \in 2^\Omega \setminus \Omega, X_i \subset X_{i+1} \\ m^\Omega(\Omega) = 1 - \sum_{i=1}^N \delta_i * a_{cqi} \\ \sum_{i=1}^N \delta_i = 1 \end{cases} \quad (11)$$

In the above equation $a_{cqi} \in [0, 1]$ is the contributor's c certainty in his i th answer X_i to question q . We have chosen to weight this certainty by a $\delta_i \in [0, 1]$ value because we do not fully trust the contributor of crowdsourcing platforms. Indeed, Gadiraju *et al.* [8] show in their article that the less competent contributors overestimate their capacity and results. Note that if the contributor does not change at all his answer X_1 when offered, then the consonant mass function proposed above becomes a simple support mass function (equation (4)) discounted (equation (5)) by a coefficient δ_1 .

Furthermore, in the event that the contributor does not know which answer to choose and selects the entire discernment frame then we have the categorical mass function $m^\Omega(\Omega) = 1$ which models total ignorance on the part of the contributor. The following section presents the results obtained from the testing of our model.

5. Experiments

We perform our experiments on real data from a crowdsourcing campaign using the 10_birds_iterative dataset from [13]. This crowdsourcing campaign focuses on the annotation of bird photos. A photo is shown to contributors with a response set Ω of 10 bird names. The contributor can choose up to 5 bird names from the set.

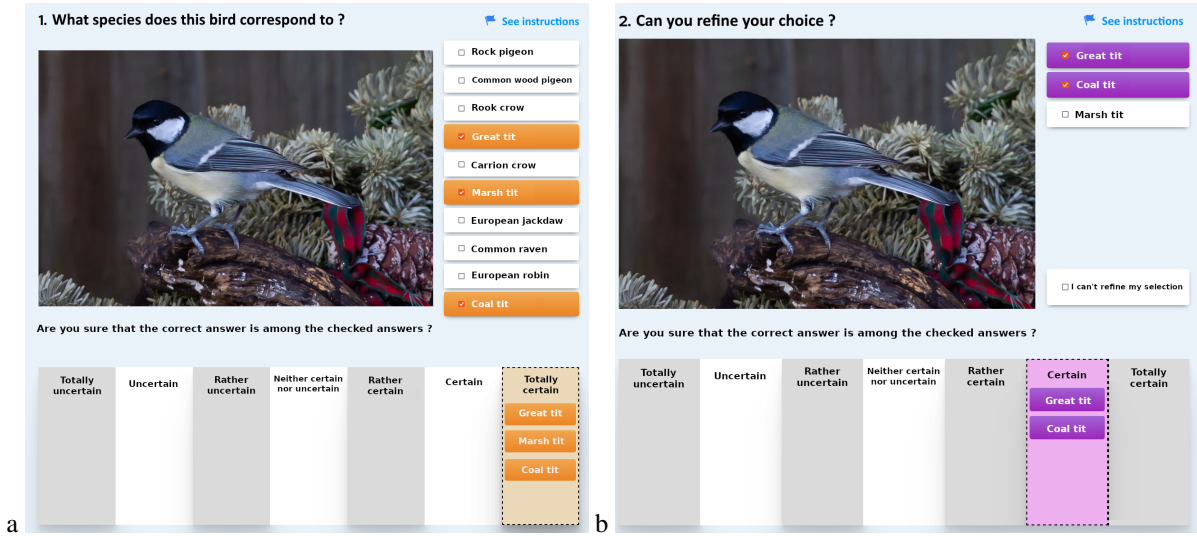


Fig. 1. Interfaces used by [13] for the campaigns 10_birds_iterative, a first iteration, b second iteration

During the 10_birds_iterative campaign, the same question q can be asked twice to the contributor c who can then extend or specify his first answer X_1 by a second answer X_2 if he wishes. The dynamic of the campaign is as follows:

- if the contributor chooses a single bird name but is not completely certain of his answer, he is then invited to broaden the selection if he wishes
- if he chooses several bird's names, he is asked to reduce his selection if he can (example given figure 1)

There is no obligation for the contributor to modify his answer and this is not penalising in any way for him.

At the time of his first selection X_1 , the contributor enters a degree of certainty of numerical value $a_{cq1} \in [0, 1]$. If he chooses to fill in a second answer X_2 he must indicate his new certainty whose numerical value is noted $a_{cq2} \in [0, 1]$.

For the 10_birds_iterative crowdsourcing campaign the crowd was composed of 51 contributors answering to 50 questions of bird annotation. 2550 responses were collected in the first round of questioning (figure 1.a) and 440 additional responses were collected in the second round (figure 1.b). For the first answer X_1 , 1417 contributions are precise ($|X_1| = 1$) and 1133 contributions are imprecise ($|X_1| > 1$), the contributor makes good use of the possibility to be imprecise in case of hesitation about the answer. In addition on the 440 answers there are:

- 88 answers for which the first answer is precise and the second one imprecise i.e. $|X_1| = 1$ and $|X_2| > 1$
- 352 answers for which the first answer is imprecise and the second one is precise $|X_1| > 1$ and $|X_1| > |X_2|$

440 of the first answers (X_1) can be used to construct consonant mass functions, the rest of the answers will be modelled by simple support mass functions.

According to the model we proposed in this article, for this data set since we have at most 2 nested responses the consonant mass function is written:

$$\begin{cases} m_{1,2}^{\Omega}(X_1) = \delta_1 * a_{cq1} \\ m_{1,2}^{\Omega}(X_2) = \delta_2 * a_{cq2} \\ m_{1,2}^{\Omega}(\Omega) = 1 - \delta_1 * a_{cq1} - \delta_2 * a_{cq2} \\ \delta_1 + \delta_2 = 1 \end{cases} \quad (12)$$

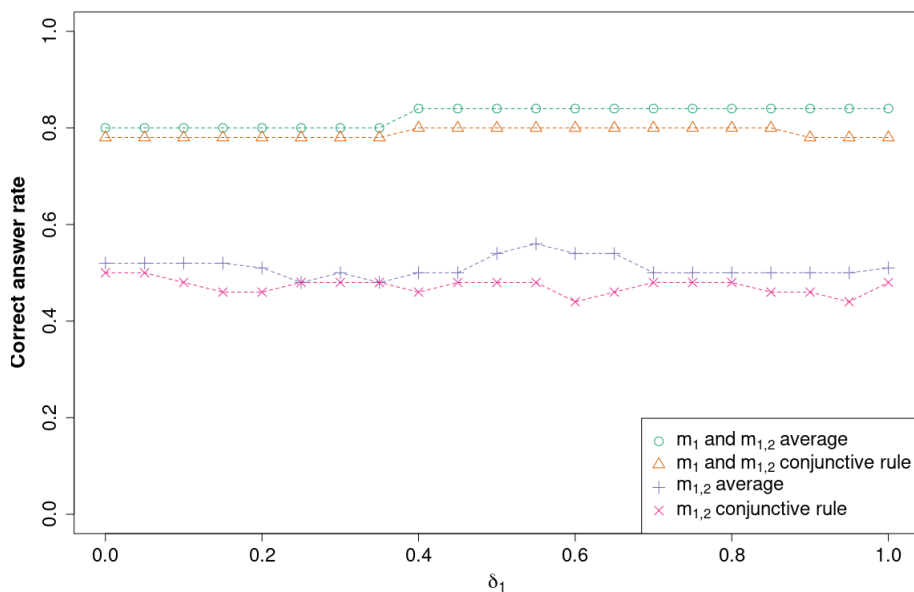


Fig. 2. Comparison of correct response rates for the full data set (2990) and the dynamic subset (440) as a function of δ_1

As mentioned in the previous section, if the contributor does not change his answer then we are in the presence of a simple support mass function. In the rest of this section, we denote m_1 the simple support mass function associated with the contributor's response X_1 in the absence of any modification, and also $m_{1,2}$ the consonant mass function associated with responses X_1 and X_2 . In total we calculate 2110 simple support mass functions (m_1) and 440 consonant mass functions ($m_{1,2}$) from the campaign data. After the aggregation we take a decision by computing the pignistic probability and select the element with the maximum of probability, which allow us to compute a correct response rate for the combination.

In Figure 2, the y-axis shows the correct response rates (CRR) of the aggregated data calculated over the 50 photos for the 2990 data by combining the responses modelled by simple support (m_1) and consonant mass functions ($m_{1,2}$). The correct response rates of the consonant mass functions alone ($m_{1,2}$) on the 440 data are also shown. The used combination operators are the normalised conjunctive operator, equation (8) and the average equation (6). The data employed for the $m_{1,2}$ functions are not employed to make other m_1 functions so that the m_1 and $m_{1,2}$ functions are independent which allows us to apply the conjunctive operator. The x-axis carries the different values of δ_1 used in the equation (11) for the computation of the consonant mass functions. The larger δ_1 is, the greater the value given to the first response X_1 from the contributor, and consequently, as $\delta_2 = 1 - \delta_1$, the value given to the second response is low. Thus for $\delta_1 = 1$ only the first response is considered for the consonant mass function so that $m_{1,2}$ is equivalent to m_1 . Note that δ_1 also has an impact on the single-support mass functions (m_1) since it acts as a discounting coefficient. Moreover for m_1 functions when $\delta_1 = 0$ then $m_1(\Omega_q) = 1$.

We observe that figure 2 shows higher rates of correct answers for the combination of the m_1 and $m_{1,2}$ functions compared to the $m_{1,2}$ function alone which is explained by the fact that we have more data in the first case than in the second. We also find that aggregation by the average gives better results than aggregation by the conjunctive operator. For the aggregation of the m_1 and $m_{1,2}$ functions, the correct response rate increases for $\delta_1 \geq 0.4$ for both operators and then decreases for $\delta_1 \geq 0.9$ for the conjunctive operator. On the other hand, the results of the aggregation of the $m_{1,2}$ functions alone do not seem to indicate an obvious correlation between δ_1 and the rate of correct response.

The maximum correct response rate obtained for all the 2990 data for the aggregation of m_1 and $m_{1,2}$ by the average is 0.84 against 0.80 for the conjunctive operator with the same functions. We also tested the case where all the 2550 responses (X_1) are modelled by a simple support mass function (m_1) and we do not take into account the second

Table 1. Comparison of the correct response rates (CRR) of m_1 only with the maximum of $m_{1,2}$ for different aggregations and datasets.

Data	Aggregation	$CRR(m_1)$	$CRR_{MAX}(m_1, m_{1,2})$
Full (2990)	Average	0.84	0.84
	Conjunctive rule	0.78	0.80
Consonant (440)	Average	0.51	0.56
	Conjunctive rule	0.48	0.50

Table 2. Number of contributor responses for different size of X_1 for the set of 440 data construction consonant mass function.

$ X_1 = 1$	$ X_1 = 2$	$ X_1 = 3$	$ X_1 = 4$	$ X_1 = 5$
88	218	118	16	0

response (X_2). In this case, we obtain a good response rate of 0.84 with the mean and 0.78 with the conjunctive operator, these results are listed in the table 1. For the average on all the data, the CCRs are in the best case equal. For the conjunctive operator, the use of consonant mass functions improves the CCR when $\delta_1 < \delta_2$, i.e. when more weight is given to the second response.

For the 440 data modelled only by the consonant mass functions $m_{1,2}$, the maximum CCR is 0.56 for the average and 0.50 for the conjunctive operator (see table 1). In the case where only the first responses (X_1) of these data are modelled by simple support mass functions then we obtain a CCR of 0.51 for the average and 0.48 for the conjunctive operator. The consonant mass functions therefore improve the CCRs.

Furthermore, as majority voting is commonly used in crowdsourcing platforms, we applied this aggregation both to the 2550 data items that constitute the first response (X_1) and also to the 440 data items that gave rise to a second response. For the integrity of the first responses we obtain a CCR of 0.76, and for the 440 responses a CCR of 0.48. We therefore obtain better results with belief function modeling of responses than with majority voting. Moreover, modeling the evolutionary responses by consonant mass functions allows to improve the good response rates obtained for an aggregation by the conjunctive operator.

Let us now turn our attention to the dissociation of the aggregation of the consonant mass functions from the first accurate and then inaccurate responses of the others. To do this, we split for the $m_{1,2}$ function curves in Figure 2 our data set into two:

- data where the contributor selects a single bird name for his first response ($|X_1| = 1$) and then eventually widens his selection ($|X_2| > |X_1|$): subset of 88 responses
- data where the contributor is initially imprecise ($|X_1| > 1$) and then eventually restricts his selection ($|X_1| > |X_2|$): subset of 352 responses

For these two sets, the responses are modelled using the functions $m_{1,2}$ by varying the value of δ_1 , the results obtained appear respectively for the average figure 3.a and for the conjunctive combination figure 3.b.

According to Figures 3 for average and conjunctive rule combination, as δ_1 increases, the CRR for first accurate then inaccurate contributions increases ($|X_1| = 1$ and $|X_2| > |X_1|$). Conversely, the CCR for first inaccurate then more accurate contributions decreases ($|X_1| > 1$ and $|X_1| > |X_2|$). It seems more interesting to give more weight to the more accurate of the two responses.

In order to observe more finely the impact of the contributor's first response, during the aggregation we will now decompose the data set for $m_{1,2}$ in order to have a subset for each possible $|X_1|$ size. We should therefore have 5 possible subsets of the data but as shown in table 2 no contributor has selected the maximum set of responses allowed i.e. 5 bird names. We can see from this table that the contributors were indeed imprecise since the contributions of the 440 data observed include mostly 2 to 3 bird names. It is not really surprising that no contributor selected 5 answers because the species names proposed in [13] campaign do not allow such hesitation.

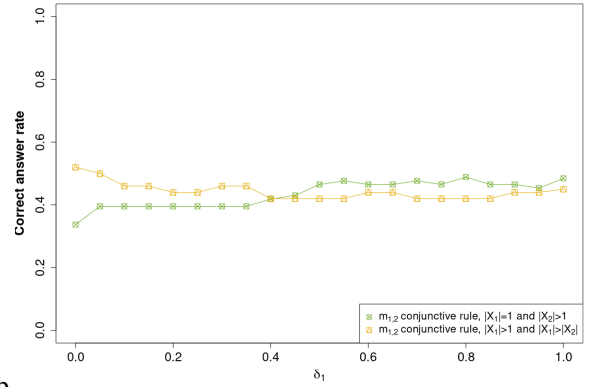
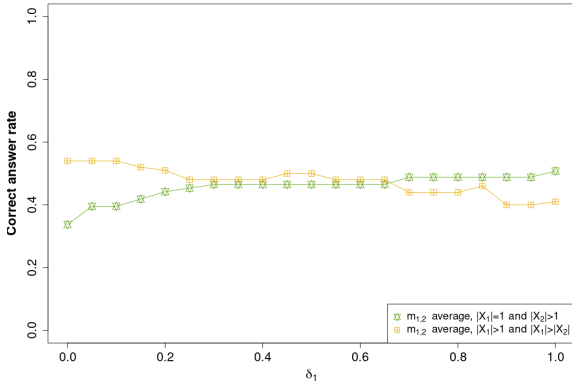


Fig. 3. Comparison of correct response rates for aggregation (a) by average, (b) by conjunctive rule of responses depending on the precision of the first (X_1) the dynamic data (440) as a function of δ_1 .

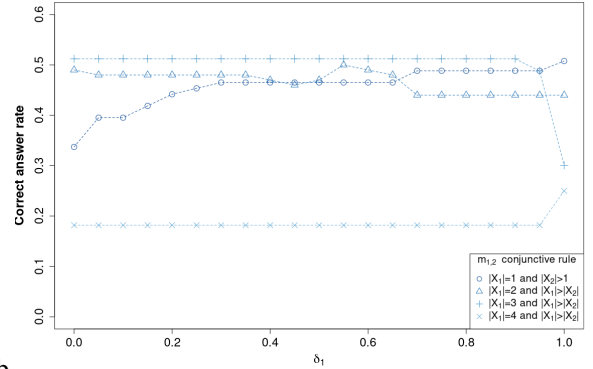
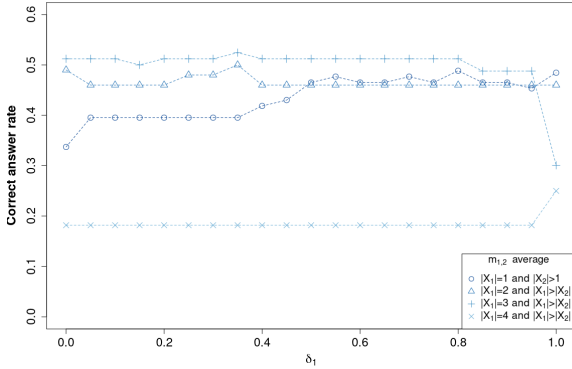


Fig. 4. Comparison of correct response rates for aggregation (a) by average, (b) by conjunctive rule of responses depending on the precision of the first (X_1) on the dynamic data (440) as a function of δ_1 .

We have 4 curves in figure 4, one for each possible size of $|X_1|$. In figure 4, the curve for which $|X_1| = 4$ is the one with the lowest rate of correct response, invariably from δ_1 . It is difficult to discuss this curve in view of the small amount of data (Table 2) that it contains. The curve for which $|X_1| = 3$ is equally constant in figure 4.b for an aggregation by the conjunctive operator until $\delta_1 > 0.9$ then the curve decreases when more weight is given to the first answer. We observe variations on figure 4.a of this curve which decreases from $\delta_1=0.8$. We also see a decrease in the $|X_1| = 2$ curve as δ_1 increases, which is more pronounced for the conjunctive operator's aggregation than for the average operator.

We therefore see from the analysis of the figures 3 and 4 that it is more relevant to favor the confidence of more accurate responses for consonant mass functions.

6. Conclusion

The tasks carried out in crowdsourcing platforms are usually in the form of multiple choice questionnaires where a single precise answer is expected from the contributor. This approach forces the contributor to select a random answer in case of doubt which is damageable for the aggregation. However, it has been shown by [14] that it is more relevant to offer the contributor the possibility of providing an imperfect answer because by being imprecise, when in doubt, he

is more certain and vice versa. In addition to this, studies [7, 6] also show that offering the contributor the possibility of changing his answer, by self-correction for example, is beneficial to him.

We propose in this paper a modelling of evolutionary responses in crowdsourcing MCQs. To do so, we consider that an evolutive answer is an answer that the contributor modifies in order to provide precision (by restricting the set of answers he has selected) or conversely by being more imprecise in order to be more certain (by selecting more answers). We choose to model the evolution of responses using consonant mass functions which are specific belief functions. We carried out our experiments to test our model on real data coming from a dynamic crowdsourcing campaign where the contributor had the possibility of changing his response. Like other studies [15, 9], our results showed that the modelling of responses and its aggregation by belief functions offers better rates of correct responses compared to majority voting. In addition, we also obtained better correct response rates through the use of consonant mass functions that model evolutionary responses compared to simple support mass functions that model non-evolutionary responses. We have also observed that the weighting applied to the responses evolution has an impact on the overall aggregation. We note that by giving more importance to more precise answers we improve the rate of correct answers.

The results obtained allow us to affirm the relevance of the modelling proposed for a consonant mass function with 3 focal elements. Indeed, the data exploited do not allow to have more than 3 focal elements, it would be interesting to carry out experiments with mass functions having more focal elements. In addition, in our future work we will question the weighting to be given to a focal element according to the size of the latter and the moment of its appearance in the evolving response of the contributor.

References

- [1] Abassi, L., Boukhris, I., 2019. A worker clustering-based approach of label aggregation under the belief function theory. *Applied Intelligence* 49, 53–62.
- [2] Dawid, A.P., Skene, A.M., 1979. Maximum likelihood estimation of observer error-rates using the em algorithm. *Applied statistics* 28, 20–28.
- [3] Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics* 38, 325–339.
- [4] Diaz, J., Rifqi, M., Bouchon-Meunier, B., 2007. QCM évidentiels pour le diagnostic des apprenants. *Actes des rencontres francophones sur la Logique Floue et ses Applications (LFA2007)*.
- [5] Diaz, J., Rifqi, M., Bouchon-Meunier, B., Jhean-Larose, S., Denhière, G., 2008. Imperfect answers in multiple choice questionnaires, in: *European Conference on Technology Enhanced Learning*, Springer. pp. 144–154.
- [6] Dow, S., Kulkarni, A., Klemmer, S., Hartmann, B., 2012. Shepherding the crowd yields better work, in: *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pp. 1013–1022.
- [7] Drapeau, R., Chilton, L., Bragg, J., Weld, D., 2016. Microtalk: Using argumentation to improve crowdsourcing accuracy, in: *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, pp. 32–41.
- [8] Gadiraju, U., Fetahu, B., Kawase, R., Siehndel, P., Dietze, S., 2017. Using worker self-assessments for competence-based pre-selection in crowdsourcing microtasks. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 1–26.
- [9] Kouloughli, D., Hadjali, A., Rassoul, I., 2016. Handling query answering in crowdsourcing systems: A belief function-based approach, in: *Fuzzy Information Processing Society (NAFIPS), 2016 Annual Conference of the North American, IEEE*. pp. 1–6.
- [10] Martin, A., 2019. Conflict management in information fusion with belief functions, in: *Information quality in information fusion and decision making*, Springer. pp. 79–97.
- [11] Shafer, G., 1976. *A mathematical theory of evidence*. volume 42. Princeton university press.
- [12] Smets, P., 1990. Constructing the pignistic probability function in a context of uncertainty. *Uncertainty in Artificial Intelligence*.
- [13] Thierry, C., Hoarau, A., Martin, A., Dubois, J.C., Le Gall, Y., 2022. Real bird dataset with imprecise and uncertain values, in: *Belief Functions: Theory and Applications: 7th International Conference, BELIEF 2022, Paris, France, October 26–28, 2022, Proceedings*, Springer. pp. 275–285.
- [14] Thierry, C., Martin, A., Dubois, J.C., Le Gall, Y., 2021. Validation of Smets’ hypothesis in the crowdsourcing environment, in: *Belief Functions: Theory and Applications: 6th International Conference, BELIEF 2021, Shanghai, China, October 15–19, 2021, Proceedings 6*, Springer. pp. 259–268.
- [15] Thierry, C., Martin, A., Dubois, J.C., Le Gall, Y., 2023. Estimation of the qualification and behavior of a contributor and aggregation of his answers in a crowdsourcing context. *Expert Systems with Applications* 216, 119496.
- [16] Yager, R.R., 1987. On the dempster-shafer framework and new combination rules. *Information sciences* 41, 93–137.