

Automated Hierarchical Conflict Reduction for Crowdsourced Annotation Tasks using Belief Functions

Constance Thierry, David Gross-Amblard, Yolande Le Gall, and Jean-Christophe Dubois

Univ Rennes, IRISA, CNRS
name.surname@irisa.fr

Abstract. A typical crowdsourcing task is concept labeling, where participants annotate e.g. images using a list of predefined concepts. Recent popular campaigns for environmental bird monitoring even use hierarchies of concepts (taxonomies of species) to obtain the most precise labeling of bird images. But in most applications, volunteer opinions are isolated from each other, and decision is taken upon majority voting. In this work we propose a new iterative labeling process where participants express their opinions together, on ascending levels of the taxonomy. Level changes are performed to minimize opinion conflict, according to the belief function theory. This complex task is orchestrated by a finite-state automaton driven by conflict measures.

Keywords: Belief functions · Crowdsourcing · Taxonomy · Automaton

1 Introduction

The crowds on crowdsourcing platforms are very diversified [16, 11]. In order to ensure the quality of the data received and to limit any noise, the employer is obliged to recruit a large number of contributors. In a classic industrial platforms (Amazon mechanical turk¹, Wirk², etc.), the employer proposes a finite number of tasks to a finite number of contributors. Some elements of the literature [1, 3, 7, 17, 22] choose to authorize or assign a task to contributors according to their level of knowledge in the field, which can limit the size of the crowd. However, this remains problematic in the event of there not being enough expert on the platform. In addition, crowdsourcing campaigns tend to be linear, and current industry platforms don't allow for a more advanced workflow. The same questions are asked of all contributors, with no dynamic processing and consequently no modulation of questions according to the difficulties encountered.

In the approach proposed here, the crowdsourcing campaign is more flexible: there's no size imposed on the crowd, and there's no finite number of tasks either. We hypothesize that the campaign enables the use of a taxonomy, such as the

¹ <https://www.mturk.com/>

² <https://www.wirk.io/>

annotation of bird photos. The idea is as follows: we allow contributors to participate in a task (with no restrictions on their expertise) as long as the aggregated responses present too significant conflict. If no agreement can be reached, the task is reformulated thanks to the taxonomy and sent out to the crowd again. The aim here is to reduce the size of the crowd required for a crowdsourcing campaign, while increasing the degree of trust placed in contributions.

Headwork³, defined by Gross-Amblard *et al.* [8], enables crowdsourcing of complex tasks by addressing a workflow with a state automaton. In this way, crowdsourcing tasks are states of the automaton, and the task workflow will depend on the automaton's rules. To achieve our goal, we are using an automaton as Headwork and a taxonomy to manage the crowdsourcing campaign workflow. A state in the automaton corresponds to a task to be carried out by the contributor where the answers given are related to a taxonomy level. The change of state is decided thanks to a conflict measure estimated by the theory of belief functions [5, 18]. As shown in previous work [19–21], this theory can be used to model contributor responses on crowdsourcing platforms. In particular, it enables contribution aggregation and decision-making. We propose to go a step further by using this theory to define when to stop asking contributors, or how to refine a question to improve the quality of the results obtained.

The contributions in this article are as follows:

- New dynamic crowdsourcing campaign model
- Workflow management of a crowdsourcing campaign by an automaton using a taxonomy
- Change of state of the automaton by computing conflict within the crowd using belief function theory

The article is structured as follows. Section 2 reviews the key elements of the belief functions required for the proposed model, while Section 3 presents the state of the art. Section 4 introduces the model and Section 5 concludes.

2 Belief functions

The section begins with a review of the basics of mass functions (Section 2.1), followed by a discussion of changes in the frame of discernment (Section 2.2).

2.1 Mass function basics

The set of classes or hypotheses r_i that are exclusive and exhaustive is the frame of discernment $\Omega = \{r_0, \dots, r_n\}$. The mass functions $m^\Omega : 2^\Omega \rightarrow [0, 1]$ model the elementary degree of belief of the source and respect the normalization condition: $\sum_{X \in 2^\Omega} m^\Omega(X) = 1$. An element $X \in 2^\Omega$ such that $m^\Omega(X) > 0$ is called focal element. If only the singletons of Ω are focal elements then m^Ω is a probability, the function is then called Bayesian mass function, it is a specific

³ <https://headwork.irisa.fr/headwork/>

mass function but there are others such as the simple support mass functions (X^ω). This mass function reflects an uncertain and imprecise response from the information source.

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

For information fusion, the sources all report on the same frame of discernment Ω . Numerous combination operators are available [14], from the average of mass functions to disjunctive and conjunctive combinations. As an example, the conjunctive rule (Equation 2) reduces the imprecision on the focal elements and increases the belief on the concordant ones.

$$m_{Conj}^\Omega(X) = \left(\bigcap_{c=1}^K m_c^\Omega \right) (X) = \sum_{Y_1 \cap \dots \cap Y_K = X} \prod_{c=1}^K m_c^\Omega(Y_c) \quad (2)$$

The mass $m_{Conj}^\Omega(\emptyset)$ represents the global conflict of the combination.

2.2 Change of discernment frame

If two sources express themselves on two different, but compatible, frames of discernment (Ω and Θ), it is possible to combine information by refining or expending one of the frames of discernment. Thanks to a refinement function $R : 2^\Omega \rightarrow 2^\Theta$:

$$m^\Theta(R(X)) = m^\Omega(X), \forall X \in 2^\Omega \quad (3)$$

For an expending of the mass function, use the reciprocal of R . However, these functions only apply to compatible discernment frames. Otherwise, it is necessary to turn to other methods. Thanks to vacuous extension and marginalization present by Delmotte *et al.* [4], it is possible to perform projections of Ω onto $\Omega \times \Theta$ and conversely. These operations are extremely useful for combination of belief functions with different discernment frameworks.

3 State of the art

Karampinas and Triantafillou [10] propose an algorithm for crowdsourcing taxonomy creation. The authors believe that humans can be useful in creating taxonomies of their knowledge domain. The objective is to aggregate a set of tags to create the taxonomy.

Farrell and Knapp [6] propose a multi-source classification based on a taxonomy using belief functions. To achieve this, the approach considers the fusion of heterogeneous input classifications that accommodate a taxonomy of one million leaf nodes with an output that can be used as a basis for decision-making. Belief functions are used to move from one level of taxonomy to another, while taking into account the possibility of opening up to the world (\emptyset). Each level of the

taxonomy $T^i = t_1^i, \dots, t_j^i$ can be interpreted as a frame of discernment ($T^i = \Omega$). T^0 is the root level and T^n the most specific level. The approach considers a taxonomy node to be the union of its child nodes. To choose the appropriate taxonomic framework for decision-making, the authors compute the information content value. We note that the authors do not use Dempster’s conjunctive combination but a Bayesian combination, and the focal elements of the mass functions are exclusively singletons.

In our approach, we propose to use the levels of a taxonomy as a frame of discernment, as Farrell and Knapp do for classification. While the authors propose a static generalist approach, we propose a dynamic model in a crowdsourcing context. Indeed, whereas the authors have a constant number of sources, our objective is to limit the number of sources and integrate more contributors only if necessary. Furthermore, we only consider contributors’ answers to the most detailed $\Omega = T^n$ discernment framework, in order to facilitate the choice of answers for non-expert contributors.

In the work of Thierry et al. [21] a crowdsourcing campaign was carried out to annotate bird photos. The identification process takes place in two stages. First, the contributor is asked to choose one or more species from the 10 proposed, and to provide a degree of certainty in his answer. Two alternatives are possible: if the contributor has chosen only one species but is not totally certain of his answer, he is then asked to complete his choice by adding other species to increase his certainty. On the other hand, if the contributor has chosen several species, he is asked whether or not he is able to narrow down his choice. A limitation of this method is that no structure is used to organize the bird species proposed. So, if the contributor hesitates during the first identification stage, the system has no information to help him make his choice.

In the next section of this article, we propose not to use the scale of the contributor but the scale of the crowd to reformulate the question we wish to ask (i.e. the contributor may not find his previous selection in the new set proposed).

4 Proposed approach

In this section, we first introduce (Section 4.1) the proposed theoretical model, then illustrate our ideas with a case study (Section 4.2).

4.1 Proposed model

Following on from previous studies[19–21], we model the responses X of contributors c to crowdsourcing Multiple Choice Questionnaire (MCQ) q using the theory of belief functions. To do this, the set of MCQ choices proposed to the contributor composes the frame of discernment Ω . The contributor indicates his certainty α in his answer X , and his contribution is modeled by a simple support mass function m_{cq}^Ω (Equation 1). The responses of all contributors are then aggregated on their frame of discernment by a combination operator to obtain

the mass function m_q^Ω . Then, a decision on the answer is made in a probabilistic framework.

In the context of crowdsourcing, the modeling of contributions using belief function theory is usually carried out with a fixed frame of discernment. We consider a crowdsourcing campaign whose response proposals can be extracted from the taxonomy T , assumed to be known at the time of campaign design. This taxonomy associated with the proposed choices is used to refine or extend the frame of discernment using the Equation 3. The campaign is directed by a deterministic state automaton. A state of the automaton corresponds to a question, address to the crowd, whose answers depend on the taxonomy T . Since the crowd has varying levels of expertise on crowdsourcing platforms, we choose as our answer set $\Omega_n = T^n$ where T^n is the taxonomy level furthest from the root and therefore the most detailed. When a minimal number of contributors, defined by the employer, have realized the question of a state, we calculate the conflict on the answers $conf(\Omega_n)$ and compared it to a threshold ϵ , then:

- If $conf(\Omega_n) < \epsilon$ then we can aggregate the crowd answers and take a decision.
- Else ($conf(\Omega_n) > \epsilon$), we compute the conflict again by changing the frame of discernment to use the values of the $l < n$ level of the taxonomy ($\Omega_l = T^l$). We calculate this new conflict until we find Ω_l such that $conf(\Omega_l) < \epsilon$. This calculation must be based on a reasonable number of levels.
 - If there is no Ω_l such as $conf(\Omega_l) < \epsilon$ then we allow new participants join the crowd and answer the question q .
 - Else ($conf(\Omega_l) < \epsilon$), we take the decision on Ω_l . Then ask the crowd the question q' which is the same as q but change the answer set to $\Omega'_n \subset \Omega_n$ such that Ω'_n is the set of leaves of the T^l node chosen. To do so, we ask contributors for whom their initial response X to q was not in the proposed answers of q' (i.e. $X \not\subseteq \Omega'_n$).

And so we continue until there is no longer any conflict over decision-making.

There are many measures of conflict in the theory of belief functions [9, 13, 15]. Among them, the value of the aggregated mass $m_q^\Omega(\emptyset)$, resulting from conjunctive rule (Equation 2), can symbolizes the conflict between the sources of information (in this case the contributors), the higher this value, the greater the conflict. So we proposed that in the following case study the transition from one state to another in the automaton depends on the value of $m_q^\Omega(\emptyset)$, which must be below a threshold value ϵ .

We assume that only responses from serious contributors are processed by the model. There are various methods for estimating the contributor's profile and seriousness [2, 7, 12, 20] that can be used to complement the model proposed in this paper.

4.2 Case study

A taxonomy of birds creates different categories to determine their relationship. Several similar bird species belong to the same genus; several similar genus belong

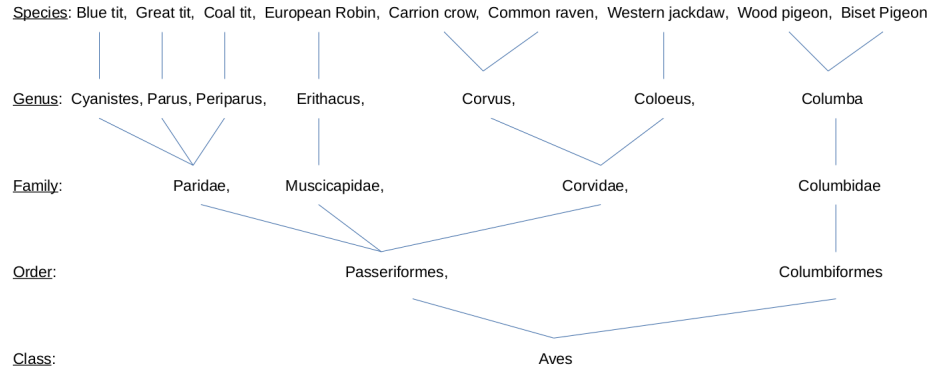


Fig. 1. Example of bird taxonomy for crowdsourcing campaign

to the same family; and several families make up an order. Taxonomists use other categories that allow more sophisticated classification, but in this case study we'll keep this simplified version. All birds together make up what is known as the bird class (Aves). In our case study, we consider a crowdsourcing campaign for bird photo annotation using the taxonomy shown in Figure 1. We therefore have the following associated frame of discernment:

- Ω_S for the **Species:** Blue tit, Great tit, Coal tit, European Robin, Carrion crow, Common raven, Western jackdaw, Wood pigeon, Biset pigeon.
- Ω_G for the **Genius:** Cyanistes, Parus, Periparus, Erithacus, Corvus, Coloeus, Columba.
- Ω_F for the **Family:** Paridae, Muscicapidae, Corvidae, Columbidae
- Ω_O for the **Order:** Passeriformes, Columbiformes

Consider a photo of a bird presented to the contributor with the question 1 (Q_1): “What species does this bird belong to?”. The proposed set of answers is composed of the elements of Ω_S . We have chosen to use the common names of the proposed species rather than the scientific names, which are more difficult for non-experts to know.

Now let's consider a crowdsourcing campaign run by a deterministic automaton whose questions will depend on the contributors' answers (Figure 2). The first question (Q_1) is open to a set of contributors with the set of answers from the frame of discernment Ω_S . There are then 3 possible scenarios.

Case 1: For the Ω_S discernment framework, the conflict on all responses is lower than the threshold value ($m^{\Omega_S}(\emptyset) < \epsilon$), so a decision can be made.

If there is too much conflict on the frame of discernment Ω_S ($m^{\Omega_S}(\emptyset) > \epsilon$), then we perform a discernment frame change for Ω_l with $l \in \{G, F, O\}$ (possible thanks to marginalization).

Case 2: If there is still too much conflict after enlarging Ω_S to Ω_l ($m^{\Omega_l}(\emptyset) > \epsilon \forall l$) then we invite new contributors to answer question Q_1 .

Case 3: After refinement, the conflict on Ω_l is acceptable. A decision can be made on Ω_l , which allows the question to be rephrased to the crowd. For example,

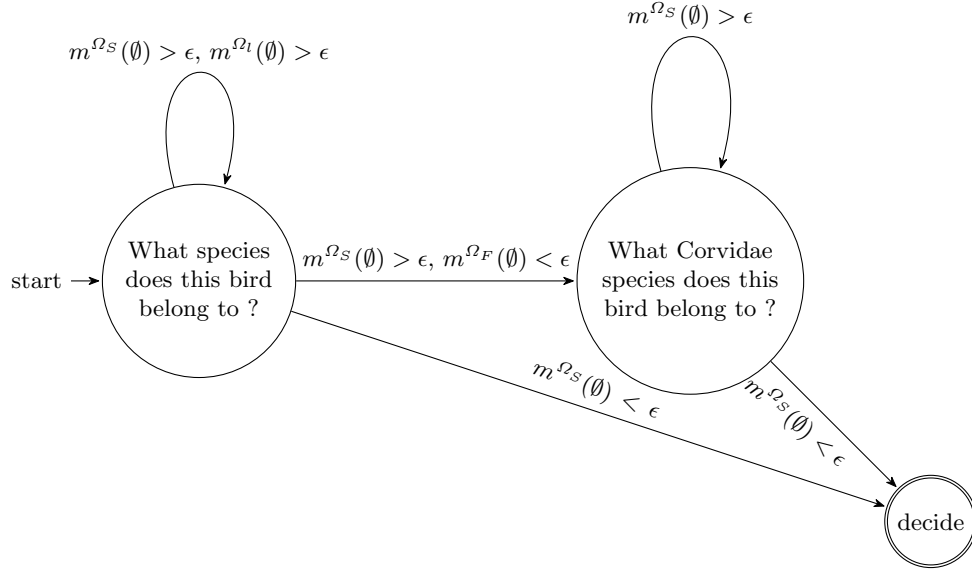


Fig. 2. Workflow with BF Example

if there is too much conflict on Ω_S and Ω_G but not on Ω_F . The decision on Ω_F indicates that the bird is a *Corvidae*. Then it's possible to ask the crowd with the same photo question 2 (Q_2): What species of *Corvidae* does this bird belong to? The discernment framework $\Omega'_S \subset \Omega_S$ used will be composed exclusively of *Corvidae* specie.

$$\Omega'_S = \{Carrion\ crow, Common\ raven, Western\ jackdaw\}$$

When in this 3rd case, we can calculate the Ω_S frame again using an expending. We check if $(m^{\Omega_S}(\emptyset) < \epsilon)$, then we proceed with the decision; otherwise, we continue questioning the crowd.

5 Conclusion and perspectives

In this paper we propose a model for dynamically managing the workflow of a crowdsourcing campaign using a taxonomy and an automaton. The automaton's state changes are decided on the basis of conflict within a crowd of contributors, calculated using belief function theory. This model will soon be integrated into the Headwork platform to be tested during real crowdsourcing campaigns and to evaluate its performance. We particularly plan to conduct a study to determine the most relevant conflict calculation in this context. Our future work also focuses on adapting the level of taxonomy to be adopted according to the profile and expertise of the contributor: more specific for experts and more general for non-experts.

References

1. Amsterdamer, Y., Davidson, S.B., Milo, T., Novgorodov, S., Somech, A.: Oassis: query driven crowd mining. In: Proceedings of the 2014 ACM SIGMOD international conference on Management of data. pp. 589–600. ACM (2014)
2. Blanco, H.H.R.: Machine-learning for spammer detection in crowd-sourcing. Human Computation AAAI Technical Report (2012)
3. Boim, R., Greenshpan, O., Milo, T., Novgorodov, S., Polyzotis, N., Tan, W.C.: Asking the right questions in crowd data sourcing. In: 2012 IEEE 28th International Conference on Data Engineering. pp. 1261–1264. IEEE (2012)
4. Delmotte, F., Smets, P.: Target identification based on the transferable belief model interpretation of dempster-shafer model. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans **34**(4), 457–471 (2004)
5. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics **38**, 325–339 (1967)
6. Farrell III, W.J., Knapp, A.M.: Multisource taxonomy-based classification using the transferable belief model. In: Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2012. vol. 8407, pp. 35–41. SPIE (2012)
7. Folorunso, O., Mustapha, O.A.: A fuzzy expert system to trust-based access control in crowdsourcing environments. Applied Computing and Informatics **11**(2), 116–129 (2015)
8. Gross-Amblard, D., Tommasi, M., Rakotoniaina, I., Thierry, C., Singh, R., Jacoboni, L.: Headwork: a data-centric crowdsourcing platform for complex tasks and participants. EDBT 2024 (2024)
9. Jousselme, A.L., Grenier, D., Bossé, É.: A new distance between two bodies of evidence. Information fusion **2**(2), 91–101 (2001)
10. Karampinas, D., Triantafyllou, P.: Crowdsourcing taxonomies. In: The Semantic Web: Research and Applications: 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012. Proceedings 9. pp. 545–559. Springer (2012)
11. Kazai, G., Kamps, J., Milic-Frayling, N.: The face of quality in crowdsourcing relevance labels: Demographics, personality and labeling accuracy (2012)
12. Khattak, F.K., Salleb-Aouissi, A.: Quality control of crowd labeling through expert evaluation. In: Proceedings of the NIPS 2nd Workshop on Computational Social Science and the Wisdom of Crowds. vol. 2, p. 5 (2011)
13. Martin, A.: About conflict in the theory of belief functions. In: Belief Functions: Theory and Applications, pp. 161–168. Springer (2012)
14. Martin, A.: Conflict management in information fusion with belief functions. In: Information quality in information fusion and decision making. pp. 79–97. Springer (2019)
15. Martin, A., Jousselme, A.L., Osswald, C.: Conflict measure for the discounting operation on belief functions. In: Information Fusion, 2008 11th International Conference on. pp. 1–8. IEEE (2008)
16. Ross, J., Zaldivar, A., Irani, L., Tomlinson, B.: Who are the turkers? worker demographics in amazon mechanical turk. Department of Informatics, University of California, Irvine, USA, Tech. Rep (2009)
17. Roy, S.B., Lykourantzou, I., Thirumuruganathan, S., Amer-Yahia, S., Das, G.: Crowds, not drones: modeling human factors in interactive crowdsourcing. In: DBCrowd 2013-VLDB Workshop on Databases and Crowdsourcing. pp. 39–42. CEUR-WS (2013)

18. Shafer, G.: A mathematical theory of evidence, vol. 42. Princeton university press (1976)
19. Thierry, C., Dubois, J.C., Le Gall, Y., Martin, A.: Modeling uncertainty and inaccuracy on data from crowdsourcing platforms: Monitor. In: Proceedings of the 31st International Conference on Tools with Artificial Intelligence (2019)
20. Thierry, C., Martin, A., Dubois, J.C., Le Gall, Y.: Estimation of the qualification and behavior of a contributor and aggregation of his answers in a crowdsourcing context. *Expert Systems with Applications* **216**, 119496 (2023)
21. Thierry, C., Martin, A., Le Gall, Y., Dubois, J.C.: Modeling evolutionary responses in crowdsourcing mcq using belief function theory. *Procedia Computer Science* **225**, 2575–2584 (2023)
22. Yongxin, T., Caleb, C.C., Chen, J.Z., Yatao, L., Lei, C.: Crowdcleaner: Data cleaning for multi-version data on the web via crowdsourcing. *Data Engineering (ICDE)*, 2014 IEEE 30th International Conference (2014)