Actes JIAF 2024

Agrégation de Jugements avec une Fiabilité Variable Inconnue

Quentin Elsaesser[1] **Patricia Everaere**[2] **Sébastien Konieczny**[1]

[1] CRIL, CNRS, Université d'Artois
[2] CRIStAL, CNRS, Université de Lille
elsaesser@cril.fr,
patricia.everaere-caillier@univ-lille.fr
konieczny@cril.fr

Résumé

Pour choisir une solution, les méthodes d'agrégation de jugements s'appuient souvent sur le nombre de votes que chaque formule reçoit de la part des agents. Cela suppose implicitement que tous les agents ont la même fiabilité et que tous les votes ont la même importance. Dans ce travail, nous considérons une vision épistémique de l'agrégation de jugements, où nous considérons qu'il existe une vérité sous-jacente. Trouver la vérité en utilisant ce que dit la majorité des agents peut conduire à une mauvaise solution, i.e. une solution où certaines formules n'ont pas la bonne valeur de vérité. L'idée dans ce travail est de suivre les opinions des agents les plus fiables pour trouver la vérité. À cette fin, nous proposons une nouvelle famille de méthodes d'agrégation de jugements qui évaluent la fiabilité des agents et des formules. Cette évaluation est ensuite utilisée pour prendre une décision et trouver la vérité, au lieu de considérer simplement le nombre de votes. Nous présentons une étude expérimentale montrant que ces méthodes donnent de meilleurs résultats pour la recherche de la vérité par rapport aux approches existantes dans la littérature. Nous étudions également les propriétés théoriques de ces méthodes.

Abstract

To choose an outcome, judgment aggregation methods often rely on the number of votes each formula receives from agents. This implicitly assumes that all agents possess equal reliability and that all votes carry identical weight. In this work we consider an epistemic view of judgment aggregation, where we consider that there is an underlying truth. Finding the truth by using what the majority of agents says can lead to a wrong solution, i.e. a solution where some of the formulae do not have the correct truth value. The idea of this work is to follow the opinions of the most reliable agents to find the truth. To this aim, we propose a new family of judgment aggregation methods that evaluate the reliability of the agents and issues. This evaluation is then used to take a decision and find the truth, instead of simply consider the number of votes. We provide an experimental study showing that these methods yield superior results in the truth-tracking task compared to existing approaches in the literature. We also study the theoretical properties of these methods.

1 Introduction

Il existe deux façons d'agréger les croyances de différents agents/sources. La première consiste à définir une nouvelle base de croyances (représentée par une formule ou un ensemble de formules) qui synthétise les croyances du groupe d'agents. Cette façon de faire correspond à la fusion des croyances [29, 19, 20, 9]. La deuxième possibilité consiste à demander à ces agents de répondre à un ensemble de questions, chaque question étant représentée par une formule logique, et chaque agent donnant une réponse oui/non en fonction du fait qu'il croit que la formule est vraie ou fausse. Ceci correspond à l'agrégation de jugements [25, 13, 22].

Même s'il existe des liens entre ces deux domaines [11, 28], ils sont principalement utiles dans des contextes différents. On ne peut donc pas choisir d'utiliser l'agrégation de jugements ou la fusion de croyances. C'est l'application qui dicte si nous avons simplement besoin d'une réponse à certaines questions ou si nous devons considérer une base de croyances complète.

Dans cet article, nous nous concentrerons sur l'agrégation de jugements. Les méthodes d'agrégation de jugements visent à former une décision collective en combinant les jugements individuels des agents qui reflètent de manière appropriée les opinions du groupe sur un ensemble de questions [25, 13].

Suivre les opinions de la majorité des agents peut conduire à des résultats incohérents. Ce dilemme, appelé *paradoxe doctrinal* ou *dilemme discursif* [21, 14, 27, 24] réside dans le fait que l'utilisation du vote majoritaire avec des agents rationnels (cohérents) peut aboutir à un résultat irrationnel (incohérent). De nombreuses règles d'agrégation ont été proposées pour résoudre ce problème, telles que les méthodes basées sur les *prémisses* [6] qui divisent l'agenda (i.e. l'ensemble des formules propositionnelles) en prémisses et en conclusions et utilisent les prémisses pour prendre une décision, les méthodes dites *séquentielles* [5] qui examinent de manière séquentielle les éléments de l'agenda, les règles basées sur les *quotas* [5] qui associent un quota à chaque proposition de l'agenda et acceptent les propositions qui dépassent ce quota. Une autre famille de règles d'agrégation de jugements utilise les distances et/ou la minimisation pour résoudre le conflit. Par exemple, les règles basées sur la *distance* [28], utilisent la distance entre les ensembles de jugements et le profil pour choisir un résultat collectif. [22] introduisent des méthodes basées sur la *minimisation*, qui minimisent la perte d'information dans le profil. Les méthodes basées sur le *soutien* [10] utilisent le soutien obtenu par les questions pour sélectionner les ensembles de jugements.

Nombre de ces méthodes utilisent le nombre de voix attribuées aux questions pour prendre une décision collective, et la plupart des méthodes, notamment celles basées sur la *minimisation* choisissent l'ensemble de jugements majoritaires (i.e. l'opinion majoritaire sur chaque question) s'il est cohérent. Mais suivre l'opinion de la majorité des agents peut conduire à une mauvaise solution d'un point de vue épistémique, même si le résultat sélectionné est cohérent. Dans l'agrégation de jugement standard, tous les agents sont considérés comme ayant la même fiabilité. Cependant, nous considérons ici le cas plus général (et plus réaliste) où les agents ont des fiabilités différentes et où la fiabilité de chaque agent est initialement inconnue. C'est ce que nous avons appelé la *fiabilité variable inconnue* dans le titre. Et notre objectif est de suivre l'opinion des agents fiables. Nous devons donc évaluer cette fiabilité et tenir compte de cette évaluation pour trouver le bon résultat.

Nous adoptons donc dans cet article une vision épistémique de l'agrégation de jugements. En général, on parle peu de la nature des jugements, qui peuvent souvent être des croyances (quelqu'un est-il coupable ou non ?) ou des objectifs/préférences (devrions-nous aller en Italie pour nos prochaines vacances ?) [1]. Dans ce travail, nous nous concentrons sur les croyances, donc sur les jugements épistémiques. Cela signifie que les agents fournissent leurs croyances dans leurs *ensembles de jugements* et qu'il existe une *vérité*, que nous aimerions identifier avec la méthode d'agrégation de jugements, comme dans la formulation originale du paradoxe doctrinal.

Illustrons notre propos par un exemple. Considérons les formules propositionnelles $\varphi_1 = p, \varphi_2 = p \wedge r, \varphi_3 = r \vee s, \varphi_4 = p \wedge q, \varphi_5 = t$. La vérité est lorsque les cinq formules

	p	$p \wedge r$	$r \vee s$	$p \wedge q$	t
A_1, A_2, A_3	1	1	1	1	1
A_4	1	0	0	1	1
A_5	1	1	1	1	0
A_6	1	1	1	0	1
A_7, A_8, A_9, A_{10}	0	0	0	0	0
A_{11}	0	0	1	0	1
Majorité	1	0	1	0	1
Vrité	1	1	1	1	1

TABLE 1 – Profile

sont vraies. La Table 1 rapporte les opinions des onze agents sur les cinq formules où 1 signifie que l'agent vote pour la formule et 0 pour la négation de la formule.

Il convient de noter que les formules de l'agenda sont généralement liées par une contrainte, i.e. que certaines formules s'impliquent mutuellement ou qu'il existe des incohérences entre certaines formules. Nous supposons que chaque agent respecte cette contrainte de cohérence (liée aux formules propositionnelles) lorsqu'il donne son avis. Une méthode de base pour prendre une décision collective consiste à utiliser le vote majoritaire sur une question donnée. Le résultat de ce vote est nommé *Majorité* dans la Table 1. Dans notre cas, l'objectif des agents est de trouver la *Vérité*. Chaque agent donne *sincèrement* son évaluation pour chaque formule, qui est la valeur de vérité de chaque formule en fonction de ses croyances. Sincèrement signifie que les agents ne sont pas censés être malveillants. Le résultat donné par le vote majoritaire est cohérent dans cet exemple, alors de nombreuses règles d'agrégation de jugements prendront ce que la majorité des agents dit comme résultat (cette propriété de sélection de l'ensemble de jugements majoritaires lorsqu'il est cohérent, appelée "Préservation de la Majorité", est souvent considérée comme une propriété attrayante). Néanmoins, ici, le vote majoritaire est différent de la vérité sur les formules φ_2 et φ_4. Les méthodes qui évaluent la fiabilité des agents et la confiance des formules à l'aide de méthodes itératives peuvent aider à trouver correctement la vérité. Dans l'exemple, les quatre agents A_7, A_8, A_9, A_{10} se trompent sur les cinq formules. Ils influencent négativement le résultat du vote majoritaire en obtenant $\neg\varphi_2$ et $\neg\varphi_4$ au lieu de la vérité φ_2 et φ_4 (i.e. $\varphi_2 = 0$ et $\varphi_4 = 0$ au lieu de $\varphi_2 = 1$ et $\varphi_4 = 1$). Si la fiabilité des agents A_7, A_8, A_9, A_{10} est correctement évaluée, la confiance dans les formules soutenues par les agents A_7, A_8, A_9, A_{10} sera faible. Inversement, la fiabilité des agents A_1, A_2, A_3 soutenant la vérité devrait être élevée et aider à trouver collectivement la vérité. L'utilisation de (l'évaluation estimée de la) confiance des formules peut aider à trouver la vérité plus efficacement qu'en se basant uniquement sur le nombre de votes que chaque question reçoit, comme le font les méthodes qui satisfont la Préservation de la Majorité.

1. Cette distinction entre *agrégation de jugements* et *agrégation de préférences* est faite au début de l'article original [21].

Dans cet article, nous définissons des méthodes qui permettent de trouver correctement la vérité sur ce genre d'exemple. Nous montrons qu'elles satisfont les propriétés logiques attendues pour les méthodes d'agrégation de jugements, et qu'elles sont plus performantes que les méthodes de la littérature pour la recherche de la vérité.

Un résultat surprenant et important de ce travail est que la propriété de Préservation de la Majorité, qui est généralement considérée comme une propriété très attrayante du point de vue de la rationalité, va à l'encontre de l'efficacité de la recherche de la vérité, comme illustré dans l'exemple ci-dessus et aussi dans nos expérimentations. La recherche de la vérité devrait être l'objectif principal des méthodes qui agrègent des croyances, ce qui signifie que les méthodes qui satisfont la Préservation de la Majorité devraient être considérées avec précaution. Plus précisément, une conséquence du théorème du jury de Condorcet (et de ses généralisations) [4, 26, 33, 12, 18] est que si nous disposons d'un nombre suffisant de sources fiables, la majorité trouvera la bonne réponse. Ainsi, dans les applications comportant des milliers de sources, la Préservation de la Majorité n'irait pas à l'encontre de la recherche de la bonne solution. Mais dans les nombreuses applications où nous ne disposons que d'un nombre relativement faible de sources, ces méthodes sont moins efficaces que les méthodes que nous proposons dans ce travail (et ces méthodes seront aussi efficaces que les méthodes satisfaisant la Préservation de la Majorité sur un grand nombre de sources).

Notons que l'étude de la recherche de la vérité avec des règles d'agrégation de jugements a été traité dans [17, 16, 1, 2]. Mais ces travaux portaient sur des structures d'agenda très particulières. [1, 2] effectue une analyse théorique des jeux en supposant un comportement stratégique des agents sur un agenda composé de 2 questions liées. [16] étudie les performances des règles basées sur les prémisses, les conclusions et la fusion sur l'agenda classique de 3 questions du paradoxe doctrinal. [17] étudie les règles basées sur les prémisses et les conclusions. Ces règles sont très particulières et nécessitent des informations supplémentaires pour savoir quelles sont les prémisses et quelles sont les conclusions, informations qui ne sont pas souvent disponibles. Leur principal objectif est d'identifier les bonnes raisons (prémisses) qui sont utilisées pour prendre la bonne décision (conclusions), et elles fournissent un résultat d'optimalité pour une classe particulière de méthodes qui satisfont une condition *d'impartialité*. Dans cet article, nous travaillons sur des agendas généraux (générés aléatoirement), et nos résultats sont donc plus robustes à cet égard. De plus, nous sommes les seuls à considérer que la fiabilité des agents peut être différente et à en tenir compte au cours du processus d'agrégation.

En ce qui concerne le plan de cet article, la section 2 présente quelques préliminaires. Nous rappelons quelques méthodes de la littérature dans la section 3. Nous expliquons comment évaluer la fiabilité des agents et la confiance des formules dans la section 4. Les *méthodes Itératives d'Évaluation de la Confiance* sont définies dans la section 5. Dans la section 6, nous vérifions quelles propriétés sont satisfaites par nos nouvelles méthodes et nous fournissons une étude expérimentale dans la section 7. La section 8 conclut.

2 Préliminaires

Cette section rappelle quelques définitions sur l'agrégation de jugement et introduit des définitions pour l'évaluation de la fiabilité.

Un *agenda* X est un ensemble fini, non vide et totalement ordonné de formules propositionnelles non triviales (i.e. non contradictoires et non tautologiques) $X = \{\varphi_1, \ldots, \varphi_m\}$. Nous désignons par $\underline{X}$ l'agenda étendu qui contient les formules de X et leur négation, i.e. $\underline{X} = \{\varphi_1, \neg\varphi_1, \ldots, \varphi_m, \neg\varphi_m\}$.

Un *jugement* sur une formule φ_k de X est un élément de $D = \{1, 0\}$, où 1 signifie que φ_k est soutenue, 0 signifie que $\neg\varphi_k$ est soutenue.

Un *ensemble de jugements* sur X est une application J de X dans D^m, qui peut également être considéré comme l'ensemble des formules $\{\varphi_k \mid \varphi_k \in X \text{ avec } J(\varphi_k) = 1\} \cup \{\neg\varphi_k \mid \varphi_k \in X \text{ avec } J(\varphi_k) = 0\}$.

Nous pouvons étendre les ensembles de jugements sur X à l'agenda étendu $\underline{X}$ directement par $J(\neg\varphi_k) = \neg J(\varphi_k)$, où $\neg J(\varphi_k) = 1$ si $J(\varphi_k) = 0$, et $\neg J(\varphi_k) = 0$ si $J(\varphi_k) = 1$.

On attend souvent des ensembles de jugements qu'ils soient cohérents et résolus : Un ensemble de jugements J sur X est *cohérent* si $\bigwedge_{\{\varphi_k \in X | J(\varphi_k)=1\}} \varphi_k \wedge \bigwedge_{\{\varphi_k \in X | J(\varphi_k)=0\}} \neg\varphi_k$ est cohérent. Il est *résolu* ssi $\forall \varphi_k \in X$, $J(\varphi_k) = 0$ ou $J(\varphi_k) = 1$.

Nous désignons par $\mathcal{A} = \{a_1, \ldots, a_n\}$ l'ensemble de tous les agents.

Un profil $P = \langle J_1, \ldots, J_n \rangle$ sur X est un vecteur d'ensembles de jugements sur X où J_i est l'ensemble de jugements de l'agent a_i. P est *cohérent* (resp. *résolu*) lorsque chaque ensemble de jugements qu'il contient est cohérent (resp. *résolu*).

Remarque. Par souci de simplification, nous ne considérerons désormais que les profils résolus et cohérents, i.e. tous les agents doivent donner leur avis (cohérent) sur toutes les formules. Nous désignons par $\mathcal{J}$ l'ensemble des ensembles de jugements cohérents et résolus.

Pour chaque formule φ_k de l'agenda étendu $\underline{X}$, nous définissons $\mathcal{A}_P(\varphi_k) = \{a_i \in \mathcal{A} \mid J_i(\varphi_k) = 1\}$ comme les agents qui soutiennent φ_k dans le profil P.

Pour un agenda X, une *méthode d'agrégation de jugements* R assigne à un profil P sur X un ensemble non vide $R(P)$ d'ensembles de jugements J sur X. Cela signifie que les méthodes d'agrégation de jugements ne doivent pas nécessairement renvoyer un résultat unique, mais tous les résultats admissibles (i.e. tous les ensembles de jugements ayant la même évaluation). Cela est nécessaire car, dans de nombreux cas, il n'existe aucun moyen de se débarrasser des liens entre certains ensembles de jugements.

3 Méthodes d'Agrégation de Jugements

Dans cette section, nous donnons quelques définitions des méthodes d'agrégation de jugements issues de la littérature. Nous rappelons les définitions et les méthodes de [22, 10].

d_H est la *distance de Hamming* : la distance de Hamming $d_H(J, J')$ entre deux ensembles de jugements J et J' est le nombre de questions sur lesquelles J et J' sont en désaccord, i.e. $d_H(J, J') = |\{\varphi_k \mid J(\varphi_k) \neq J'(\varphi_k)\}|$. La distance de Hamming $d_H(P, P')$ entre deux profils est $d_H(P, P') = \sum_{i=1}^{n} d_H(J_i, J_i')$.

$N_P(\varphi) = |\mathcal{A}_P(\varphi)|$ est le nombre d'agents dans P dont l'ensemble de jugements comprend φ.

Tout sous-ensemble cohérent de l'agenda peut être étendu afin d'obtenir un ensemble de jugement résolu. Pour tout $S \subseteq X$ cohérent, l'ensemble des extensions rationnelles de S est $ext(S) = \{J \mid J \in \mathcal{J}$ et $S \subseteq J\}$.

L'ensemble de jugements majoritaires associé à un profil P contient tous les éléments de l'agenda qui sont soutenus par une majorité d'ensembles de jugements dans P : $m(P) = \{\varphi \in X \mid N_P(\varphi) > \frac{n}{2}\}$. Un profil P est majoritairement cohérent si $m(P)$ est cohérent.

Étant donné un ensemble de formule S et une formule φ, $S' \subseteq S$ est dit φ-cohérent si $S' \cup \{\varphi\}$ est cohérent. S' est un sous-ensemble maximal φ-cohérent de S si S' est φ-cohérent et s'il n'existe pas de $S'' \subseteq S$ qui soit φ-cohérent et que $S' \subset S''$. $max(S, \subseteq)$ désigne les sous-ensembles maximaux cohérent de S. L'ensemble $S' \subseteq S$ est un sous-ensemble *maxcard* φ-cohérent de S si S' est φ-cohérent et qu'il n'existe pas d'ensemble φ-cohérent $S'' \subseteq S$ tel que $|S'| < |S''|$. Le sous-ensemble cohérent *maxcard* de S est noté $max(S, |.|)$.

Il existe deux méthodes basées sur des ensembles de jugements cohérents maximaux : La règle utilisant les *sous-ensembles d'agenda maximaux* R_{MSA} et la règle utilisant les *sous-ensembles d'agenda maxcard* R_{MCSA}. Elles sont définies dans [22] comme suit :

Définition 1. *Pour tout agenda X, pour tout profil P,* $R_{MSA}(P) = \{ext(S) \mid S \in \max(m(P), \subseteq)\}$ *et* $R_{MCSA}(P) = \{ext(S) \mid S \in \max(m(P), |.|)\}$.

La règle basée sur l'ensemble des jugements majoritaires est appelée règle du *sous-agenda à poids maximal* R_{MWA} et est définie comme suit :

Définition 2. $R_{MWA}(P) = \operatorname{argmax}_{J \in \mathcal{J}} \sum_{\varphi \in J} N_P(\varphi)$.

Une méthode inspirée de la règle de vote "avec rangement de paires" [30], nommé R_{RA}, est définie comme suit :

Définition 3. *Soit* $\geq_P$ *le pré-ordre total sur* $\underline{X}$ *définie par* $\varphi \geq_P \phi$ *si* $N_P(\varphi) \geq N_P(\phi)$. R_{RA} *est défini par la procédure algorithmique suivante :*

réorganiser les éléments de $\underline{X}$ *de telle sorte que* $\varphi_1 \geq_P \ldots \geq_P \varphi_{2m}$
$D := \emptyset$
pour $k = 1, \ldots, 2m$ *faire*
si $D \cup \{\varphi_k\}$ *est cohérent* ***alors***
$D \leftarrow D \cup \{\varphi_k\}$
fin si
fin pour
$R_{RA}(P) := D.$

La règle basée sur la distance $R^{d_H,max}$ est définie comme suit :

Définition 4. $R^{d_H,max}(P) = \operatorname{argmin}_{J \in \mathcal{J}} \max_{i=1}^{n}(d_H(J_i, J))$

Enfin, nous rappelons la définition des méthodes basées sur le soutien $\delta^{RM\oplus}$ venant de [10].

Un ensemble de jugements $J = \{\varphi_1, \ldots, \varphi_m\}$, avec $\varphi_i \in \underline{X}$, est évalué en termes de soutien de ses éléments. Plus précisément, $s(J)$ est le vecteur de score associé à l'ensemble de jugement J tel que $s(J) = (N_P(\varphi_1), \ldots, N_P(\varphi_m))$.

Une méthode d'agrégation par jugement majoritaire ordonnée $\delta^{RM\oplus}$ (où $\oplus$ peut être $leximax$, $leximin$ ou Σ) est définie comme suit :

Définition 5. *Une méthode d'agrégation par jugement majoritaire ordonnée* $\delta^{RM\oplus}$ *associe à chaque profil P sur l'agenda X l'ensemble de jugements J avec le vecteur de score le plus élevé s par rapport à la fonction d'agrégation* $\oplus$, *i.e.* $\delta^{RM\oplus}(P) = \{J \in \mathcal{J} \mid \nexists J' \in \mathcal{J} t.q. \oplus(s(J')) > \oplus(s(J))\}$.

4 Évaluation de la confiance

La plupart des méthodes de la littérature sur l'agrégation de jugements utilisent le nombre de votes reçus par chaque formule pour établir un jugement collectif et considèrent que tous les agents ont la même fiabilité.

Ici, nous considérons que les agents ont des fiabilités différentes et que ces fiabilités sont inconnues. Cette hypothèse semble plus réaliste dans de nombreux contextes.

Pour avoir un meilleur jugement, nous voulons utiliser une méthode qui calcule la fiabilité des agents, représentée par $r(a_i) \in [0, 1]$ et la confiance de la formule $\varphi_k \in \underline{X}$, représentée par $c(\varphi_k) \in \mathbb{R}^+$. Nous utilisons une méthode itérative pour évaluer la fiabilité des agents. Au lieu de voter simplement pour la formule (ou sa négation) que l'agent estime correcte, nous mettons à jour le poids (i.e. la fiabilité) des agents à chaque itération et calculons la confiance des formules et des négations.

L'évaluation de la fiabilité est basée sur l'algorithme de découverte de la vérité proposé dans [7] (d'autres algorithmes de découverte de la vérité peuvent être trouvés dans [15, 32, 31]).

Dans ces travaux, les méthodes de découverte de la vérité considèrent des faits, objets et sources. Les objets sont plus ou moins des questions ; les faits sont les réponses possibles à ces questions (liées à une seule question) ; et les sources sont des agents qui donnent la réponse qu'ils choisissent aux questions (au plus une par objet). Dans notre cas, les formules sont les objets et les faits sont la valeur de vérité attribuée à chaque formule (0 si l'agent soutient la négation de la formule, i.e. $\neg\varphi$ et 1 si l'agent soutient la formule, i.e. φ). Un agent donne une évaluation pour toutes les formules de l'agenda, le profil est donc résolu.

Nous normalisons la fiabilité d'un agent par le nombre de formules dans l'agenda X car les profils sont résolus. L'étape de normalisation nous permet de considérer le poids associé à chaque agent comme une probabilité, qui représente la confiance empirique que nous pouvons avoir dans cet agent. Plus la fiabilité est proche de 1, plus nous considérons l'agent comme fiable.

La méthode originale est simplifiée dans le cas de l'agrégation de jugements car il n'y a que deux faits par objet (les deux valeurs possibles pour une formule) et le résultat des deux normalisations proposées dans l'article original est le même avec les profils résolus.

Nous décrivons maintenant la procédure d'évaluation de la fiabilité des agents et de la confiance dans les formules. Nous initialisons la fiabilité des agents à 1 car nous n'avons aucune information *a priori* sur les agents et nous considérons donc qu'ils sont tous *a priori* aussi fiables les uns que les autres.

Une itération se déroule comme suit : Nous commençons par évaluer la confiance des formules. Pour chaque formule, nous additionnons la fiabilité des agents qui soutiennent cette formule :

$$c(\varphi_k) = \sum_{a_i \in \mathcal{A}_P(\varphi_k)} r(a_i)$$

Ensuite, pour chaque formule, nous comparons la confiance de la formule positive $c(\varphi_k)$ et la confiance de sa négation $c(\neg\varphi_k)$. Celle qui a la plus grande confiance donne un score de 1 aux agents qui contiennent cette formule dans leur ensemble de jugement individuel et de 0 pour la formule qui a le moins de confiance. En cas d'égalité, la formule et sa négation donnent un score de 0,5 aux agents :

$$V(\varphi_k) = \begin{cases} 1 & si\ c(\varphi_k) > c(\neg\varphi_k) \\ 0 & si\ c(\varphi_k) < c(\neg\varphi_k) \\ 0.5 & si\ c(\varphi_k) = c(\neg\varphi_k) \end{cases}$$

Ensuite, pour calculer la fiabilité d'un agent, nous agrégeons les scores $V(\varphi_k)$ et nous normalisons la fiabilité de chaque agent par le nombre de formules m pour nous assurer que nous avons une fiabilité entre 0 et 1 :

$$r(a_i) = \frac{\sum_{\{\varphi_k | J_i(\varphi_k)=1\}} V(\varphi_k)}{|X|}$$

Algorithme 1 Évaluation de la confiance

Entrée : Le profil P et l'agenda X.
Sortie : La fiabilité r des agents et la confiance c des formules.

$r(a_i) = 1\ \forall a_i \in \mathcal{A}$ #fiabilité initiales des agents
$c(\varphi_k) = 0\ \forall \varphi_k \in \underline{X}$ #confiance initiale des formules
ta (resp. ta^{-1}) est le vecteur de fiabilité des agents pour l'itération courante (resp. précédente), i.e. $ta = \langle r(a_i) : \forall a_i \in \mathcal{A} \rangle$.
tant que distance_euclidean(ta, ta^{-1}) > 0.001 et
nombre_d'itérations < 30 **faire**
  # Évaluation de la fiabilité des formules
  pour chaque $\varphi_k \in \underline{X}$ **faire**
    $c(\varphi_k) = \sum_{a_i \in \mathcal{A}_P(\varphi_k)} r(a_i)$
  fin pour
  # Évaluation du score V pour chaque formule
  pour chaque $\varphi_k \in \underline{X}$ **faire**
    si $c(\varphi_k) > c(\neg\varphi_k)$ **alors**
      $V(\varphi_k) = 1$
    sinon si $c(\varphi_k) < c(\neg\varphi_k)$ **alors**
      $V(\varphi_k) = 0$
    sinon si $c(\varphi_k) = c(\neg\varphi_k)$ **alors**
      $V(\varphi_k) = 0.5$
    fin si
  fin pour
  pour chaque $a_i \in \mathcal{A}$ **faire**
    # Évaluation de la fiabilité des agents
    $r(a_i) = \frac{\sum_{\{\varphi_k | J_i(\varphi_k)=1\}} V(\varphi_k)}{|X|}$
  fin pour
fin tant que

L'algorithme (voir Algorithme 1) s'arrête après 30 itérations ou lorsque le processus converge, i.e. lorsque la distance euclidienne entre la fiabilité des agents de l'itération précédente et l'itération actuelle est inférieure à ϵ avec $\epsilon = 0.001$ [2]. Lorsque l'algorithme s'arrête, la confiance des formules devient le support des formules (la somme de la fiabilité des agents qui soutiennent les formules).

Une idée naturelle est de considérer l'ensemble de jugements obtenu en sélectionnant une formule ou sa négation en fonction de la meilleure confiance à la fin du calcul (ou les ensembles de jugements en cas d'égalité). Cet ensemble est représenté par l'ensemble $H(P)$:

Définition 6. *Nous désignons par $H(P)$ l'ensemble des ensembles de jugements cohérents dont les élé-*

2. Dans les expériences de la section 7, le nombre d'itérations n'excède jamais 7.

ments ont une confiance plus élevée que leur négation : $H(P) = \{\langle\varphi_1, \ldots, \varphi_m\rangle \mid \bigwedge \varphi_i \nvdash \perp \textit{ et } \forall k, c(\varphi_k) \geq c(\neg\varphi_k),\ \textit{avec } \varphi \in \underline{X}\}$.

Notez que la définition de $H(P)$ est faite après le calcul de la confiance de l'algorithme 1, et que $H(P)$ peut être vide si jamais la conjonction des formules n'est pas cohérente. Si $H(P)$ n'est pas vide, alors $H(P)$ sera le résultat de la méthode d'agrégation utilisant la confiance. Si $H(P)$ est vide, nous choisirons le meilleur ensemble de jugements possible, en fonction de la fiabilité des formules qu'il contient. La signification de « meilleur » conduit à des choix différents, selon des stratégies différentes. Ces stratégies seront expliquées dans la section suivante.

5 Méthodes Itératives d'Évaluation de la Confiance

Nous proposons une nouvelle famille de méthodes d'agrégation de jugements basées sur l'évaluation de la fiabilité des agents et de la confiance des formules, appelées méthodes *ICE* pour Méthodes Itératives d'Évaluation de la Confiance (*Iterated Confidence Evaluation Methods* en anglais).

Dans notre contexte, il n'existe qu'une seule *vraie* solution . Cette solution fait partie de l'ensemble de jugements $\mathcal{J}$. Les agents ne savent pas quelle solution est la vérité et l'objectif des méthodes ICE est de la trouver. A cette fin, nous définissons un vecteur de confiances (calculé avec la méthode détaillée dans la section 4) pour tous les ensembles de jugements.

Définition 7. *Étant donné l'agenda étendu* $\underline{X}$ *et un profil* P *sur* $\underline{X}$, *nous désignons par* $s(J) = \langle c(\varphi_k) \mid \varphi_k \in J\rangle$ *le vecteur de fiabilité pour un ensemble de jugements* $J \in \mathcal{J}$.

Une méthode d'agrégation de jugements utilisant la confiance avec la fonction d'agrégation $\oplus$ est définie comme suit :

Définition 8. *Étant donné un profil* P *sur un agenda* X. *Soit* $R^{\oplus}(P)$ *le(s) ensemble(s) de jugements dans* $\mathcal{J}$ *qui maximise(nt) la fiabilité des formules :*

$$R^{\oplus}(P) = \{J_1 \in \mathcal{J} \mid \oplus s(J_1) \geq \oplus s(J_2)\ \forall J_2 \in \mathcal{J}\}$$

Il est important de souligner que si $H(P)$ n'est pas vide, quelle que soit la fonction d'agrégation $\oplus$ considérée, les méthodes ICE sélectionneront $H(P)$ (voir la propriété de cohérence de la fiabilité dans la section 6).

Pour définir les méthodes ICE, nous utilisons $\Sigma, \times$ et *leximax* comme fonctions d'agrégation.

R^{Σ} maximise la somme de la fiabilité de chaque formule, ce qui est lié à une maximisation de la fiabilité globale. $R^{\times}$ maximise le produit de la fiabilité des formules. Le produit est un moyen naturel de calculer la probabilité d'un ensemble d'événements. R^{lex} consiste à trier, dans un ordre décroissant, la fiabilité des formules et à choisir la meilleure (maximum) pour l'ordre lexicographique (*leximax*). C'est une façon de maximiser la meilleure formule (du point de vue de la fiabilité) dans les ensembles de jugements sélectionnés.

Notons que le *leximin* (qui maximise la pire formule) n'est pas utilisé ici, car il a été prouvé dans [10] que le résultat de *leximax* et de *leximin* est le même pour un profil résolu. Nous écrivons *lex* au lieu de *leximax* et *leximin*.

Illustrons le comportement de nos méthodes sur l'exemple de l'introduction, et montrons que, contrairement aux autres méthodes de la littérature, elles parviennent à trouver la vérité.

Exemple 1. *Considérons le profil P de la Table 1. Nous évaluons tout d'abord la fiabilité. Nous pouvons voir l'évolution à chaque itération de la fiabilité des agents dans la Table 2 et l'évolution de la confiance des formules dans la Table 3. L'algorithme s'arrête après 4 itérations dans cet exemple. On constate que notre algorithme évalue correctement la fiabilité des agents. Les agents qui affirment la vérité* (A_1, A_2, A_3) *ont la plus grande fiabilité alors que les agents qui se trompent complètement* (A_7, A_8, A_9, A_{10}) *ont la plus faible confiance et n'influencent donc pas négativement la décision.*

	A_1, A_2, A_3	A_4	A_5	A_6	A_7, A_8, A_9, A_{10}	A_{11}
It1	1	1	1	1	1	1
It2	0.6	0.6	0.4	0.8	0.4	0.8
It3	0.7	0.5	0.5	0.9	0.3	0.7
It4	1	0.6	0.8	0.8	0	0.4

Table 2 – Fiabilité des agents

	φ_1	$\neg\varphi_1$	φ_2	$\neg\varphi_2$	φ_3	$\neg\varphi_3$	φ_4	$\neg\varphi_4$	φ_5	$\neg\varphi_5$
It1	6	5	5	6	6	5	5	6	6	5
It2	3.6	2.4	3	3	3.8	2.2	2.8	3.2	4	2
It3	4	1.9	3.5	2.4	4.2	1.7	3.1	2.8	4.2	1.7
It4	5.2	0.4	4.6	1	5	.6	4.4	1.2	4.8	0.8

Table 3 – Confiance des formules

Nous pouvons maintenant utiliser les méthodes ICE. La solution proposée est : $R^{\Sigma}(P) = R^{\times}(P) = R^{lex}(P) = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5\}$ *ce qui correspond à la vérité. Pour cet ensemble de jugements, le vecteur de fiabilité est* $s(\{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5\}) = \langle 5.2, 4.6, 5, 4.4, 4.8\rangle$ *Le score de* $R^{\Sigma}(P)$ *est 24, le score de* $R^{\times}(P)$ *est 2525.95 et pour* $R^{lex}(P)$ *dans l'ordre nous avons :* φ_1, φ_3, φ_5, φ_2 *puis* φ_4.

La plupart des méthodes de la littérature ($R_{RA}, R_{MWA}, R_{MCSA}, R_{MSA}$, $\delta^{RM_{leximax}}$ *et* $\delta^{RM_{\Sigma}}$) *satisfont la Préservation de la Majorité, alors ils donneront l'ensemble de jugements majoritaires comme solution (i.e.* $\{\varphi_1, \neg\varphi_2, \varphi_3, \neg\varphi_4, \varphi_5\}$*), ce qui est faux.* $R^{d_H,max}$ *ne satisfait pas à la Préservation de la Majorité et donne comme solutions* $\{\{\varphi_1, \neg\varphi_2, \varphi_3, \varphi_4, \neg\varphi_5\}, \{\varphi_1, \neg\varphi_2, \varphi_3, \neg\varphi_4, \varphi_5\}, \{\varphi_1, \neg\varphi_2, \varphi_3, \neg\varphi_4, \neg\varphi_5\}, \{\varphi_1, \neg\varphi_2, \neg\varphi_3, \varphi_4, \varphi_5\}\}$*, qui sont quatre solutions erronés.*

6 Propriétés

Dans cette section, nous étudions les propriétés théoriques des méthodes ICE $R^{\oplus}$.

La première propriété est *Universalité* et stipule que tous les profils cohérents doivent être une solution possible.

Universalité (Uni.). Le domaine de R est l'ensemble de tous les profils cohérents.

Proposition 1. *Les méthodes $R^{\oplus}$ satisfont* ***Universalité****.*

Démonstration. Les solutions de R sont choisis parmi les ensembles de jugements $\mathcal{J}$. Tous les ensembles de jugements de $\mathcal{J}$ sont cohérents. □

La propriété suivante est *Anonymat*. Nous voulons que les méthodes d'agrégation de jugements soient impartiales et ne favorisent aucun agent en particulier.

Anonymat (Ano.). Pour deux profils $P = (J_1, \ldots, J_n)$ et $P' = (J'_1, \ldots, J'_n)$ dans le domaine de R qui sont des permutations l'un de l'autre, nous avons $R(P) = R(P')$.

Proposition 2. *Les méthodes $R^{\oplus}$ satisfont* ***Anonymat****.*

Démonstration. La fiabilité d'un agent est évaluée en fonction de ses affirmations et n'est pas liée à l'ordre donné dans le profil. Dans notre algorithme (Algorithme 1), la fiabilité des agents sera la même en prenant le profil P ou le profil P' avec la permutation, alors les solutions données par les méthodes ICE sont les mêmes : $R(P) = R(P')$. □

Pour la propriété suivante, nous devons rappeler la *règle d'agrégation majoritaire R^{maj}* qui est définie comme suit : Pour tout agenda X et tout profil P sur X, nous avons $R^{maj}(P) = \{R_P^{maj}\}$, où pour toute question $\varphi_k \in X$:

$$R_P^{maj}(\varphi_k) = \begin{cases} 1 & \text{si } N_P(\varphi_k) > \frac{|\mathcal{A}|}{2} \\ 0 & \text{si } N_P(\neg\varphi_k) > \frac{|\mathcal{A}|}{2} \\ \star & \text{sinon} \end{cases}$$

La propriété suivante est la *Préservation de la Majorité*. Elle stipule que le résultat du vote majoritaire doit être la solution proposée par la méthode si le résultat du vote majoritaire est cohérent et résolu.

Préservation de la Majorité (Maj.). Si R_P^{maj} est cohérent et résolu alors $R(P) = \{R^{maj}(P)\}$

Proposition 3. *Les méthodes $R^{\oplus}$ ne satisfont pas* ***Préservation de la Majorité****.*

Démonstration. Considérons le profil de la Table 1, l'évaluation de la fiabilité des agents de la Table 2 et la confiance des formules de la Table 3.

Dans cet exemple, l'ensemble de jugements majoritaires cohérents est $(\{\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4, \neg\varphi_5\})$, mais les méthodes ICE donnent $\{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5\}$ comme solution. Le score de $\{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5\}$ maximise le score pour R^{Σ} et $R^{\times}$. R^{lex} ordonne les vecteurs de confiance pour obtenir le maximum $\{5.2, 5, 4.8, 4.6, 4.4\}$ (dans un ordre décroissant) et sélectionne ensuite $\{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5\}$.

Soulignons que dans ce cas, une méthode qui satisfait Préservation de la Majorité ne trouvera pas la vérité, contrairement à la nôtre. □

Si nous considérons l'exemple de la Table 1, nous voyons que suivre les opinions de la majorité des agents peut conduire à une mauvaise solution. Cet exemple montre donc que cette propriété n'est pas adéquate dans ce cadre de fiabilité à variable inconnue, où nous ne voulons pas nous concentrer sur le plus grand nombre de votes, mais sur la plus grande confiance (le plus grand nombre de votes provenant de sources fiables). Par conséquent, nous ne voulons pas que nos méthodes ICE satisfassent cette propriété.

Les méthodes ICE ne satisfont pas *Préservation de la Majorité*, mais lorsque $H(P)$ est cohérent (l'ensemble des ensembles de jugements cohérents dont les éléments ont une confiance plus élevée que leur négation), le résultat des méthodes ICE sera les formules de $H(P)$. Nous définissons une nouvelle propriété appelée *Cohérence de la fiabilité*. Cette propriété garantit qu'une méthode sélectionne les résultats cohérents les plus fiables, si possible.

Cohérence de la fiabilité (Coh.). Si $H(P)$ n'est pas vide, alors la solution de R est $H(P)$, i.e. $R(P) = H(P)$.

Proposition 4. *Les méthodes $R^{\oplus}$ satisfont* ***Cohérence de la fiabilité****.*

Démonstration. Supposons que $H(P)$ n'est pas vide.

Soit $J^R \in H(P)$. Par définition de $H(P)$, $\forall \varphi_k \in J^R, c(\varphi_k) \geq c(\neg\varphi_k)$. Supposons que $R^{\oplus}$ ne sélectionne pas J^R. Cela signifie que $\exists J \notin J^R$ t.q. $\oplus_{\varphi_k \in J}(c(\varphi_k)) > \oplus_{\varphi'_k \in J^R}(c(\varphi'_k))$. Nécessairement, cela implique que pour au moins un k, $c(\varphi_k) > c(\varphi'_k)$, i.e que $c(\neg\varphi'_k) > c(\varphi'_k)$ avec $\varphi'_k \in J^R$: contredit la définition de $H(P)$. Donc $H(P) \subseteq R^{\oplus}(P)$.

Soit $J \in R^{\oplus}(P)$. Supposons que $J \notin H(P)$. Comme $J \notin H(P)$, il y a au moins un k t.q. $c(\varphi_k) < c(\neg\varphi_k)$. Comme $H(P)$ n'est pas vide, il existe au moins un ensemble de jugement $J^R \in H(P)$ t.q. $\neg\varphi_k \in J^R$. Nous savons également que $\forall \varphi_i \in J^R, i \neq k, c(\varphi_i) \geq c(\neg\varphi_i)$. Alors $\oplus_{\varphi_i \in J}(c(\varphi_i)) < \oplus_{\varphi'_i \in J^R}(c(\varphi'_i))$: J ne peut être sélectionné par $R^{\oplus}$. Nous avons aussi une contradiction et $R^{\oplus}(P) \subseteq H(P)$.

Enfin, $R^{\oplus}(P) = H(P)$. □

Unanimité est la propriété suivante. Lorsqu'une formule est présente dans tous les ensembles de jugements individuels, elle doit l'être dans tous les résultats.

Unanimité (Una.). Pour tout $\varphi_k \in X$, si $J_i(\varphi_k) = x$ avec $x \in \{0, 1\}$, $\forall J_i \in P$, alors pour tout $J \in R(P)$, nous avons $J(\varphi_k) = x$.

Proposition 5. *$R^{\times}$ et R^{lex} satisfont* ***Unanimité****. R^{Σ} ne satisfait pas* ***Unanimité****.*

Avec R^{Σ}, la combinaison de la fiabilité de la négation de la formule unanime avec la fiabilité des formules qui maximisent leur confiance par rapport à leur négation peut dans certains cas compenser la confiance nulle de la négation de la formule unanime.

Démonstration. Considérons φ_k la formule assignée à chaque ensemble de jugement individuel et $\neg\varphi_k$ la formule qui n'est dans aucun ensemble de jugement.

Nous savons que la confiance de φ_k sera la plus élevée car tous les agents affirment que cette formule est la vérité ($c(\neg\varphi_k) = \sum_{a_i \in \mathcal{A}_P(\varphi_k)} r(a_i)$) et que la confiance en φ_k sera égale à zéro ($c(\neg\varphi_k) = \sum_{a_i \in \emptyset} 0$). De plus, nous savons qu'il y a au moins un ensemble de jugements qui contient φ_k car φ_k est dans chaque ensemble de jugements individuel.

Tous les vecteurs de fiabilité maximale pour lex contiennent φ_k. R^{lex} choisira son résultat dans ces ensembles. Nous savons que φ_k a le niveau de confiance le plus élevé et que φ_k figure au moins dans un ensemble de jugement (cohérent). Donc φ_k sera dans tous les résultats de R^{lex}.

Tous les ensembles de jugement qui contiennent $\neg\varphi_k$ obtiennent une confiance de 0 avec $R^{\times}$. Il existe au moins un ensemble de jugements cohérents, disons J, qui contient φ_k, car φ_k appartient aux ensembles de jugements de tous les agents, qui sont cohérents. Comme il y a unanimité sur φ_k, tous les agents ont une fiabilité strictement supérieure à 0. Alors toutes les formules de J ont une confiance strictement supérieure à 0. En conséquence, tout ensemble de jugement sélectionné par $R^{\times}$ contiendra φ_k.

Donc $R^{\times}$ et R^{lex} satisfont *Unanimité*.

R^{Σ} ne satisfait pas *Unanimité*. Considérons les formules propositionnelles $\varphi_1 = a$, $\varphi_2 = b$, $\varphi_3 = c$, $\varphi_4 = d$, $\varphi_5 = \neg a$, $\varphi_6 = \neg b$, $\varphi_7 = \neg c$, $\varphi_8 = \neg d$, $\varphi_9 = a \vee b$, $\varphi_{10} = \neg(a \wedge b \wedge c \wedge d)$. Nous avons quatre agents dans le profil P, avec les agents qui affirment $\{\varphi_1, \varphi_2, \varphi_3, \neg\varphi_4, \neg\varphi_5, \neg\varphi_6, \neg\varphi_7, \varphi_8, \varphi_9, \varphi_{10}\}$, $\{\varphi_1, \varphi_2, \neg\varphi_3, \varphi_4, \neg\varphi_5, \neg\varphi_6, \varphi_7, \neg\varphi_8, \varphi_9, \varphi_{10}\}$, $\{\varphi_1, \neg\varphi_2, \varphi_3, \varphi_4, \neg\varphi_5, \varphi_6, \neg\varphi_7, \neg\varphi_8, \varphi_9, \varphi_{10}\}$ et $\{\neg\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \neg\varphi_6, \neg\varphi_7, \neg\varphi_8, \varphi_9, \varphi_{10}\}$.

φ_1	$\neg\varphi_1$	φ_2	$\neg\varphi_2$	φ_3	$\neg\varphi_3$	φ_4	$\neg\varphi_4$	φ_5	$\neg\varphi_5$
2.4	0.8	2.4	0.8	2.4	0.8	2.4	0.8	0.8	2.4
φ_6	$\neg\varphi_6$	φ_7	$\neg\varphi_7$	φ_8	$\neg\varphi_8$	φ_9	$\neg\varphi_9$	φ_{10}	$\neg\varphi_{10}$
0.8	2.4	0.8	2.4	0.8	2.4	3.2	0	3.2	0

TABLE 4 – Confiance des formules du profil P

Nous avons la confiance des formules dans la Table 4. Les solutions proposées par R^{Σ} sont $R^{\Sigma} = \{ \{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \neg\varphi_5, \neg\varphi_6, \neg\varphi_7, \neg\varphi_8, \varphi_9, \neg\varphi_{10}\}, \{\varphi_1, \varphi_2, \varphi_3, \neg\varphi_4, \neg\varphi_5, \neg\varphi_6, \neg\varphi_7, \varphi_8, \varphi_9, \varphi_{10}\}, \{\varphi_1, \varphi_2, \neg\varphi_3, \varphi_4, \neg\varphi_5, \neg\varphi_6, \varphi_7, \neg\varphi_8, \varphi_9, \varphi_{10}\}, \{\varphi_1, \neg\varphi_2, \varphi_3, \varphi_4, \neg\varphi_5, \varphi_6, \neg\varphi_7, \neg\varphi_8, \varphi_9, \varphi_{10}\}, \{\neg\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \neg\varphi_6, \neg\varphi_7, \neg\varphi_8, \varphi_9, \varphi_{10}\} \}$

Les quatre solutions ont un score de 22,4 avec Σ. Les agents sont unanimes sur φ_{10} mais dans l'une des solutions de R^{Σ} nous avons $\neg\varphi_{10}$. Donc la méthode R^{Σ} ne satisfait pas *Unanimité*. □

Maintenant nous parlons de *Systématicité*, une propriété critiquée dans la littérature [3, 8, 10, 23]. Elle stipule que toutes les formules ayant le même nombre de voix doivent être traitées de la même manière. La principale critique de cette propriété est qu'elle ignore le reste du profil lorsqu'une décision doit être prise pour une formule. C'est déjà un problème dans l'agrégation standard des jugements. Mais c'est encore plus un problème dans notre cadre de fiabilité variable inconnue.

Systématicité (Syst.). Pour deux profils quelconques $P = (J_1, \ldots, J_n)$ et $P' = (J'_1, \ldots, J'_n)$ dans le domaine de R, et deux propositions quelconques φ_k et φ_l de X, tel que $J_i(\varphi_k) = J'_i(\varphi_l) \forall i$, si $J_P(\varphi_k) = x$ pour tous $J_P \in R(P)$, alors $J'_{P'}(\varphi_l) = x$ pour tous $J'_{P'} \in R(P')$.

Proposition 6. *Les méthodes $R^{\oplus}$ ne satisfont pas* ***Systematicity****.*

Cette propriété ne tient pas compte des autres formules. Mais l'avis des agents sur toutes les formules doit être pris en compte pour trouver le bon résultat. C'est ainsi que la fiabilité des agents est estimée.

Démonstration. Considérerons le profil P et P' respectivement de la Table 5 et de la Table 6 avec $\varphi_1 = p, \varphi_2 = p \wedge r, \varphi_3 = r \vee s, \varphi_4 = p \wedge q, \varphi_5 = t$. Considérerons $k = 3$ et $l = 4$, les agents expriment donc le même avis sur φ_3 et φ_4. Les agents A_6 et A_7 changent leurs opinions sur φ_1 et l'agent A_{10} sur φ_5.

Nous avons la confiance des formules dans la Table 7 pour le profil P et dans la Table 8 pour le profil P'.

	φ_1	φ_2	φ_3	φ_4	φ_5
A_1	1	1	1	1	1
A_2	1	1	1	1	1
A_3	1	1	1	1	0
A_4	1	0	0	1	1
A_5	1	0	0	1	0
A_6	0	0	1	0	0
A_7	0	0	1	0	1
A_8	0	0	0	0	0
A_9	0	0	0	0	0
A_{10}	0	0	0	0	0

TABLE 5 – Profile P

	φ_1	φ_2	φ_3	φ_4	φ_5
A_1	1	1	1	1	1
A_2	1	1	1	1	1
A_3	1	1	1	1	0
A_4	1	0	0	1	1
A_5	1	0	0	1	0
A_6	1	0	1	0	0
A_7	1	0	1	0	1
A_8	0	0	0	0	0
A_9	0	0	0	0	0
A_{10}	0	0	0	0	1

TABLE 6 – Profile P'

φ_1	$\neg\varphi_1$	φ_2	$\neg\varphi_2$	φ_3	$\neg\varphi_3$	φ_4	$\neg\varphi_4$	φ_5	$\neg\varphi_5$
1.2	4.4	0.2	5.4	1.6	4	1.2	4.4	1	4.6

TABLE 7 – Confiance des formules du profil P

φ_1	$\neg\varphi_1$	φ_2	$\neg\varphi_2$	φ_3	$\neg\varphi_3$	φ_4	$\neg\varphi_4$	φ_5	$\neg\varphi_5$
5	0.8	2.2	3.6	3.6	2.2	3.6	2.2	3.6	2.2

TABLE 8 – Confiance des formules du profil P'

Avec le profil P, la solution de nos méthodes avec Σ, $\times$ et lex est la même qu'avec le profil P et P' et est $R^{\oplus}(P) =$

$\{\neg\varphi_1, \neg\varphi_2, \neg\varphi_3, \neg\varphi_4, \neg\varphi_5\}$. Avec le profil P', nous avons $R^{\Sigma}(P') = \{\varphi_1, \neg\varphi_2, \varphi_3, \varphi_4, \varphi_5\}$, qui est différent sur φ_3 et φ_4 malgré le fait que les opinions des agents n'ont pas changé. Nous avons $R(P) \neq R(P')$ même si $\forall i, J_i(\varphi_k) = J'_i(\varphi_l)$.

□

La propriété suivante est *Neutralité*. Les éléments de l'agenda doivent être considérés de la même manière.

Neutralité (Neut.). Si $X = \{\varphi_1, \ldots, \varphi_m\}$ et $X' = \{\varphi'_1, \ldots, \varphi'_m\}$ sont deux agendas tels qu'il existe une permutation σ sur $\{1, \ldots, m\}$ satisfaisant $\varphi_k \equiv \varphi'_{\sigma(k)}$ pour tous $k \in \{1, \ldots, m\}$ alors pour tout profil $P = (J_1, \ldots, J_n)$ sur X et $P' = (J'_1, \ldots, J'_n)$ sur X' tel que pour chaque $i \in \{1, \ldots, n\}$, pour tous $k \in \{1, \ldots, m\}$, $J_i(\varphi_k) = J'_i(\varphi'_{\sigma(k)})$, nous avons $R(P) = R(P')$.

Proposition 7. *Les méthodes $R^{\oplus}$ satisfont* ***Neutralité****.*

Démonstration. Une permutation sur X ne changera pas les ensembles de jugements dans P de sorte que la confiance des formules sera la même avec P' parce que la confiance n'est pas évaluée en fonction de l'ordre des formules dans l'agenda. Une permutation sur X ne changera pas les résultats de R : $R(P) = R(P')$.

Nos méthodes satisfont à la règle *Neutralité*. □

Les résultats sont résumé dans la Table 9.

	Uni.	Ano.	Maj.	Coh.	Una.	Syst.	Neut.
R^{Σ}	✓	✓	✗	✓	✗	✗	✓
$R^{\times}$	✓	✓	✗	✓	✓	✗	✓
R^{lex}	✓	✓	✗	✓	✓	✗	✓

TABLE 9 – Propriétés satisfaites par les méthodes ICE

Les méthodes $R^{\oplus}$ satisfont donc les propriétés importantes et standard tel que Universalité, Anonymat, Neutralité et Unanimité, à l'exception de R^{Σ} qui ne satisfait pas à Unanimité. Cette méthode est donc peut-être moins intéressante que les autres pour cette raison. Aucune de ces méthodes ne satisfait Systématicité, qui est une propriété très critiquée. Elles ne satisfont pas à la propriété de Préservation de la Majorité, qui est généralement considérée comme une propriété souhaitable, mais nous avons montré que dans ce cadre, suivre la majorité des votes n'est pas toujours approprié. Nous développerons davantage ce point dans la section suivante, où nous montrerons les meilleures performances de nos méthodes en matière de recherche de la vérité par rapport à celles qui satisfont la Préservation de la Majorité.

7 Expérimentations

Nous avons procédé à une évaluation expérimentale de nos méthodes ICE afin de tester leurs performances en matière de recherche de vérité, i.e. de savoir si elles trouvent correctement la vérité pour chaque formule. [3]

À notre connaissance, il n'existe pas de jeux de données disponibles pour la recherche de la vérité dans l'agrégation de jugements. L'utilisation de jeux de données tels que ceux venant de *Preflib* [4] n'est pas possible pour plusieurs raisons. Tout d'abord, ces jeux de données concernent les préférences et sont utiles pour tester les règles de vote. La deuxième raison est qu'il n'y a pas de vérité dans Preflib, car il s'agit de préférences et non de jugements épistémiques (l'objectif de nos méthodes est de trouver la vérité). Même si nous pouvons traduire les profils de préférences en profils d'agrégation de jugements, nous ne pouvons pas utiliser ces jeux de données pour évaluer la fiabilité moyenne des profils, ce qui est un paramètre important dans nos expériences. Pour ces raisons, nous générons des agendas et des profils aléatoires afin de tester les performances des méthodes.

Pour tester toutes les méthodes, nous générons l'agenda, composé d'un ensemble de formules propositionnelles. Dans cet article, nous considérons un langage basé sur 10 propositions atomiques et chaque agenda est composé de $m = 10$ formules propositionnelles.

Pour créer aléatoirement un agenda X, nous générons chaque formule une par une. Pour générer une formule propositionnelle φ de X, nous sélectionnons aléatoirement une distance d et une interprétation ω. Pour chaque interprétation ω', nous disons que ω' est un modèle de φ si la distance de Hamming entre ω et ω' est inférieure à d : $d_H(\omega, \omega') \leq d$.

Une fois l'agenda fixé, nous pouvons choisir la vérité et générer le profil. Pour s'assurer que l'ensemble des jugements d'un agent est résolu et cohérent, nous choisissons au hasard l'un des ensembles de jugements dans $\mathcal{J}$ pour être son ensemble de jugements. La vérité, noté T, est également choisie au hasard dans $\mathcal{J}$.

Avec l'ensemble des jugements des agents, nous pouvons calculer la fiabilité moyenne des agents, i.e. la probabilité *a posteriori* de trouver les valeurs correctes pour toutes les formules en comparant l'ensemble des jugements de tous les agents et la vérité T. La fiabilité moyenne d'un agent $J_i \in P$ est $\frac{|T \cap J_i|}{|T|}$.

Chaque méthode produit un ensemble de solutions, et différentes mesures peuvent être utilisées pour estimer à quel point cet ensemble de solutions est proche de la vérité. Nous avons choisi celle qui est directement liée à la prise de décision, i.e. le nombre de décisions qui peuvent être prises à l'unanimité sur chaque formule du profil. Ainsi, toutes les solutions doivent avoir la même valeur pour une formule pour qu'il y ait une décision unanime sur cette formule. Ensuite, nous évaluons ces décisions par rapport à la vérité T choisie précédemment.

3. Le code et les agendas utilisés dans les expériences sont disponibles sur `https://github.com/QuentinElsaesser/JA`.

4. `www.preflib.org`

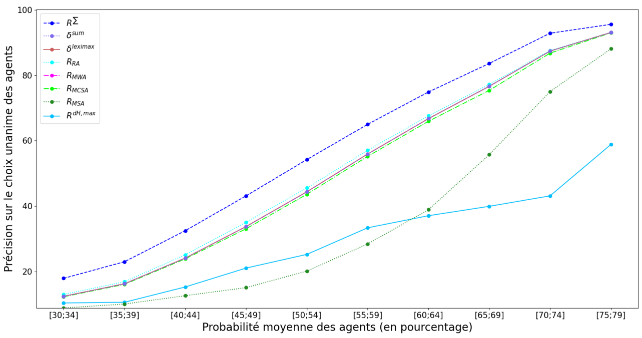

FIGURE 1 – Précision - 10 agents

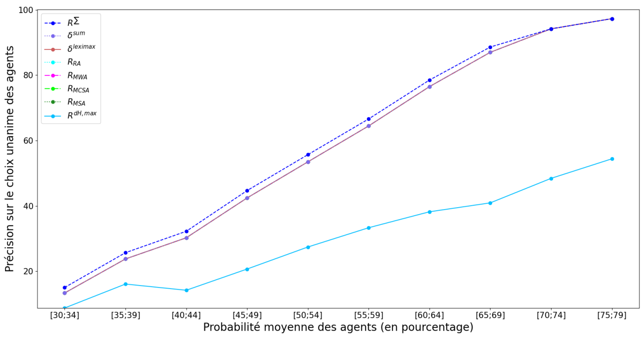

FIGURE 2 – Précision - 11 agents - Ensemble de jugements majoritaires

Définition 9. *Nous désignons par* $U(P)$ *l'ensemble des formules qui se trouvent dans tous les ensemble de jugement de* $R(P)$: $U(P) = \{\varphi_k \in \underline{X} \mid \varphi_k \in \bigcap R(P)\}$

Pour évaluer la performance, nous regardons la proportion de valeurs correctement prédites par la méthode et nous l'appelons *Précision* i.e. $Precision = \frac{|\{\varphi_k \in U(P) \mid \varphi_k \in T\}|}{|X|}$.

Chaque point sur les graphiques est la moyenne obtenue avec la génération de 1000 expériences. Nous classons les expériences en fonction de la fiabilité moyenne des agents.

Nous comparons les méthodes ICE avec les méthodes R_{MSA}, R_{MCSA}, R_{MWA}, R_{RA}, $R^{d_H,max}$ dans [22] et les méthodes de [10] $\delta^{RM_{leximax}}$ et $\delta^{RM_{\Sigma}}$. Nous ne présentons que les résultats de R^{Σ} (et non ceux de $R^{\times}$ et R^{lex}) car les solutions proposées par les trois méthodes ICE sont identiques à plus de 95% dans nos expérimentations.

Pour les expériences, nous nous concentrons sur deux paramètres qui ont un impact sur les résultats des méthodes. Le premier est la cohérence ou l'incohérence de l'ensemble de jugements majoritaires. S'il est cohérent, la plupart des méthodes de la littérature donneront l'opinion majoritaire comme résultat. Mais que se passe-t-il lorsque l'ensemble des jugements majoritaires n'est pas cohérent pour les méthodes de la littérature et comment les méthodes ICE traitent-elles ces deux cas différents ? Le deuxième paramètre est le nombre d'agents, plus précisément si le nombre d'agents est pair ou impair. Dans la littérature, le nombre d'agents est souvent impair car cela garantit une majorité d'opinions sur toutes les formules, alors que dans l'autre cas, il peut y avoir des égalités (i.e. le même soutien entre la formule et sa négation), ce qui rend les méthodes de la littérature moins décisives.

Sur la Figure 1, nous avons les résultats de la *Précision* avec 10 agents sous une fiabilité moyenne variable. Nous avons entre 10 et 50% de profils pour lesquels l'ensemble des jugements majoritaires est cohérent. Ce cas est le moins optimal pour les méthodes de la littérature. Les méthodes ICE surpassent les méthodes de la littérature de plus de 5% et lorsque la fiabilité moyenne est supérieure à 75%, les méthodes ICE sont 3% plus efficaces et sont proches de la vérité, i.e. que la *Précision* est proche de 100%. La raison pour laquelle les méthodes de ICE sont plus performantes que les méthodes de la littérature est qu'elles ne proposent en moyenne qu'une seule solution (entre 1,01 et 1,055 en fonction de la fiabilité moyenne des agents dans nos expériences) alors que les autres méthodes proposent plusieurs solutions, entre 2 et 3 et pour $R^{d_H,max}$ plus de 8. Les méthodes ICE sont donc plus décisives que les méthodes de la littérature lorsque la plupart des ensembles de jugements majoritaires sont incohérents. L'une des raisons de ce manque de décision est que les méthodes de la littérature ne savent pas comment gérer correctement les égalités entre une formule et sa négation, et donnent deux solutions possibles, l'une contenant la formule et l'autre sa négation (notons que nous n'avons pas d'égalités sur tous les profils dans cette expérience). Leur *Précision* sera donc plus faible car les résultats ne sont pas unanimes pour toutes les formules. Grâce à l'évaluation itérée de la fiabilité, les méthodes ICE permettent une meilleure gestion des égalités et donne de meilleurs résultats.

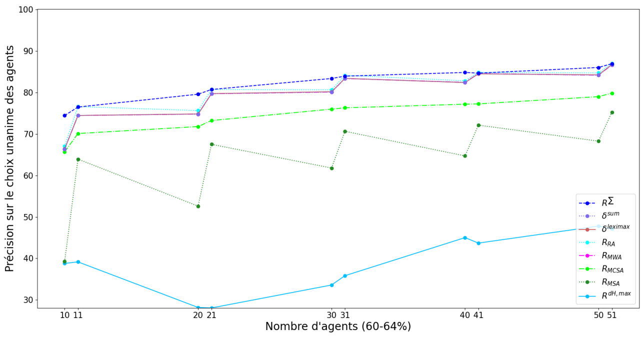

FIGURE 3 – Précision

Dans la Figure 2, le nombre d'agents est impair, de sorte qu'il n'y a plus de cas avec une égalité. De plus, nous ne sélectionnons que les profils ayant un ensemble de jugements majoritaires cohérents pour cette expérience, car il s'agit du cas le plus optimal pour les méthodes de la littérature, ils donnent l'ensemble de jugements majoritaires et sont décisifs. Nous voulons donc voir si le caractère décisif des méthodes ICE est la seule raison des bons résultats de la Figure 1. Et nous voyons que la Précision entre les méthodes ICE et les méthodes de la littérature est plus proche. Pour les méthodes ICE, il n'y a pas de changement significatif dans les résultats, et les méthodes ICE ont toujours de meilleurs résultats. Dans ces expérimentations, l'opinion d'une majorité d'agents fiables, lorsqu'elle est cohérente, est souvent proche de la vérité.

Dans la Figure 3, nous voyons les résultats dans le cas général et augmentons le nombre d'agents pour les profils avec une fiabilité fixe, entre 60 et 64%. Nous examinons le cas avec un nombre pair et impair d'agents. Dans cette expérience, plus de 76% des profils ont un ensemble de jugements majoritaires cohérents avec un nombre impair d'agents et moins de 50% avec un nombre pair d'agents. Avec un ensemble de jugements majoritaires cohérents, les méthodes ICE sont légèrement plus performantes que les méthodes de la littérature, mais ont plus de 5% avec un ensemble de jugements majoritaires incohérents. Avec un plus grand nombre d'agents, les performances de toutes les méthodes augmentent.

Nous avons observé différents cas. Lorsque l'ensemble des jugements majoritaires est cohérent, ce qui est le cas optimal pour les méthodes d'agrégation de jugements dans la littérature, les méthodes ICE sont légèrement meilleures. Cependant, lorsque l'ensemble de jugements majoritaires n'est pas cohérent, les méthodes ICE sont plus performantes. Elles sont plus décisives que les autres méthodes et peuvent traiter des cas plus complexes tels que des profils où la majorité conduit à des égalités. En moyenne, les méthodes ICE obtiennent donc de meilleurs résultats que les méthodes de la littérature pour la recherche de la vérité avec différents paramètres. L'évaluation de la fiabilité a un réel impact pour prendre la bonne décision lorsque le nombre d'agents est pair et lorsqu'il y a égalité entre le support d'une formule et sa négation.

8 Conclusion

Dans cet article, nous présentons un cadre pour l'agrégation de jugements où la fiabilité de l'agent est variable mais inconnue. Dans ce cadre, il est possible d'estimer cette fiabilité en confrontant les ensembles de jugements des agents. Nous présentons la famille des méthodes ICE, qui utilisent la confiance des formules au lieu du nombre de votes pour trouver le résultat correct. Nous montrons que ces méthodes satisfont les propriétés standard des méthodes d'agrégation de jugements et nous effectuons une évaluation montrant qu'elles sont plus performantes que les méthodes de la littérature pour la tâche de recherche de la vérité. Il est intéressant de noter qu'avec un nombre suffisant d'agents fiables, la majorité donnera la vérité (ce qui peut être considéré comme une conséquence du théorème du Jury de Condorcet), de sorte que les méthodes de la littérature qui satisfont la propriété de la Préservation de la Majorité donneront de bons résultats. Mais nous discutons du fait que la Préservation de la Majorité va également à l'encontre de la recherche de la vérité dans certains cas (les moins décisifs). Et nos méthodes ICE (qui ne satisfont pas la propriété Préservation de la Majorité) sont plus performantes que les autres méthodes lorsque cette majorité n'est pas suffisamment claire. En particulier, nous avons montré que pour un nombre pair d'agents, ou lorsqu'il n'y a pas d'ensemble de jugements majoritaires cohérent, les méthodes ICE sont nettement plus performantes.

Pour les travaux futurs, il serait possible de commencer l'évaluation de la fiabilité des agents à partir d'une fiabilité *a priori* donnée, qui pourrait provenir par exemple de la réputation de chaque agent, ou d'autres types de mesures de confiance, puis d'ajuster ces fiabilités *a priori* à l'aide de nos méthodes.

Remerciements

Ce travail a bénéficié du support de la Chaire IA BE4musIA (ANR-20-CHIA-0028).

Références

[1] Irem Bozbay. Truth-tracking judgment aggregation over interconnected issues. *Social Choice and Welfare*, 53 :337–370, 2019.

[2] Irem Bozbay, Franz Dietrich, and Hans Peters. Judgment aggregation in search for the truth. *Games and Economic Behavior*, 87 :571–590, 2014.

[3] Bruce Chapman. Rational aggregation. *politics, philosophy & economics*, 1(3) :337–354, 2002.

[4] Marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* Imprimerie royale Paris, 1785.

[5] Franz Dietrich and Christian List. Judgment aggregation by quota rules. *Journal of Theoretical Politics*, 2005.

[6] Franz Dietrich and Philippe Mongin. The premiss-based approach to judgment aggregation. *Journal of Economic Theory*, 145(2) :562–582, 2010.

[7] Quentin Elsaesser, Patricia Everaere, and Sébastien Konieczny. Voting-based methods for evaluating sources and facts reliability. In *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 178–185, 2023.

[8] Ulle Endriss, Umberto Grandi, and Daniele Porello. Complexity of judgment aggregation. *Journal of Artificial Intelligence Research*, 45 :481–514, 2012.

[9] Patricia Everaere, Chouaib Fellah, Sébastien Konieczny, and Ramón Pino Pérez. Weighted merging of propositional belief bases. In *Proceedings of the*

20th International Conference on Principles of Knowledge Representation and Reasoning, KR'23, Rhodes, Greece, September 2-8, 2023, pages 219–228, 2023.

[10] Patricia Everaere, Sébastien Konieczny, and Pierre Marquis. Counting votes for aggregating judgments. In Ana L. C. Bazzan, Michael N. Huhns, Alessio Lomuscio, and Paul Scerri, editors, *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14, Paris, France, May 5-9, 2014*, pages 1177–1184. IFAAMAS/ACM, 2014.

[11] Patricia Everaere, Sébastien Konieczny, and Pierre Marquis. Belief merging versus judgment aggregation. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS'15, Istanbul, Turkey, May 4-8, 2015*, pages 999–1007, 2015.

[12] Patricia Everaere, Sébastien Konieczny, and Pierre Marquis. The epistemic view of belief merging : can we track the truth ? In *Nineteenth European Conference on Artificial Intelligence (ECAI'10)*, pages 621–626, 2010.

[13] Davide Grossi and Gabriella Pigozzi. Introduction to judgment aggregation. In *European Summer School in Logic, Language and Information*, volume 7388 of *Lecture Notes in Computer Science*, pages 160–209. Springer, 2010.

[14] G. Th. Guilbaud. Theories of the general interest, and the logical problem of aggregation. *Readings in Mathematical Social Science*, pages 262–307, 1966.

[15] Manish Gupta and Jiawei Han. Heterogeneous network-based trust analysis : a survey. *SIGKDD Explorations Newsletter*, 13(1) :54–71, 2011.

[16] Stephan Hartmann and Gabriella Pigozzi. Judgment aggregation and the problem of truth-tracking. In *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK XI). S*, volume 248, pages 248–252, 2007.

[17] Stephan Hartmann and Jan Sprenger. Judgment aggregation and the problem of tracking the truth. *Synthese*, 187 :209–221, 2012.

[18] Patrick Hummel. Jury theorems with multiple alternatives. *Social Choice and Welfare*, 34(1) :65–103, 2010.

[19] Sébastien Konieczny and Ramón Pino Pérez. Merging information under constraints : a logical framework. *Journal of Logic and Computation*, 12(5) :773–808, 2002.

[20] Sébastien Konieczny and Ramón Pino Pérez. Logic based merging. *Journal of Philosophical Logic*, 40(2) :239–270, 2011.

[21] Lewis A Kornhauser and Lawrence G Sager. Unpacking the court. *Yale Law Journal*, 96 :82, 1986.

[22] Jérôme Lang, Gabriella Pigozzi, Marija Slavkovik, and Leendert W. N. van der Torre. Judgment aggregation rules based on minimization. In Krzysztof R. Apt, editor, *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2011), Groningen, The Netherlands, July 12-14, 2011*, pages 238–246. ACM, 2011.

[23] Jérôme Lang, Marija Slavkovik, and Srdjan Vesic. Agenda separability in judgment aggregation. In Dale Schuurmans and Michael P. Wellman, editors, *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 1016–1022. AAAI Press, 2016.

[24] Christian List. The discursive dilemma and public reason. *Ethics*, 116(2) :362–402, 2006.

[25] Christian List. The theory of judgment aggregation : an introductory review. *Synthese*, 187(1) :179–207, 2012.

[26] Christian List and Robert E. Goodin. Epistemic democracy : Generalizing the Condorcet jury theorem. *Journal of Political Philosophy*, 9(3) :277–306, 2001.

[27] Christian List and Clemens Puppe. Judgement aggregation : A survey. In Paul Anand, Prasanta K. Pattanaik, and Clemens Puppe, editors, *The Handbook of Rational and Social Choice - an overview of new foundations and applications*, pages 457–482. Oxford University Press, 2009.

[28] Gabriella Pigozzi. Belief merging and the discursive dilemma : an argument-based account to paradoxes of judgment aggregation. *Synthese*, 152 :285–298, 2006.

[29] P. Z. Revesz. On the semantics of arbitration. *International Journal of Algebra and Computation*, 7(2) :133–160, 1997.

[30] Markus Schulze. A new monotonic and clone-independent single-winner election method. *Voting Matters*, 17 :9–19, 2003.

[31] Joseph Singleton and Richard Booth. Towards an axiomatic approach to truth discovery. *Journal of Autonomous Agents and Multi-Agent Systems*, 36(2) :1–49, 2022.

[32] Dalia Attia Waguih and Laure Berti-Equille. Truth discovery algorithms : An experimental evaluation. *arXiv preprint arXiv :1409.6428*, 2014.

[33] Hobart Peyton Young and Arthur Levenglick. A consistent extension of Condorcet's election principle. *SIAM Journal on Applied Mathematics*, 35(2) :285–300, 1978.