RÉPUBLIQUE DE CÔTE D'IVOIRE

Union-Discipline-Travail

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Institut National Polytechnique
Félix HOUPHOUËT-BOIGNY

EDP
École Doctorale Polytechnique

THÈSE

Pour l'obtention du grade de

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE FÉLIX HOUPHOUËT-BOIGNY

Mention : **Sciences et Technologies**

Spécialité : **Informatique**

SUJET :

MODÉLISATION DE LA QUALITÉ DE L'INFORMATION DANS L'INTERNET DES OBJETS

Présentée et soutenue publiquement le 13 décembre 2019 par

M. MAMBE Digrais Moïse

JURY

M. BABRI Michel	Professeur Titulaire	INP-HB	Président
M. BEDJA Koffi-Sa	Professeur Titulaire	ENSI Lomé	Rapporteur
M. AKA Boko	Professeur Titulaire	UNA	Rapporteur
M. DIBY Kadjo Ambroise	Maître de Conférences	UFHB	Examinateur
M. OUMTANAGA Souleymane	Professeur Titulaire	INP-HB	Directeur de thèse

Résumé

Dans cette thèse, nous abordons le problème de la qualité de l'information dans l'Internet des Objets, un nouveau paradigme connu aussi sous l'appellation objets connectés. L'un des principaux enjeux de ce concept consiste à intégrer de l'intelligence dans les systèmes ou applications afin de les rendre autonomes. Il ne s'agit plus pour le système de passer directement à l'action lorsqu'il obtient une information, mais de mener une étape de vérification avant toute prise de décision. En effet, les informations issues des capteurs peuvent être sujettes à de nombreuses imperfections pouvant entraver la prise de décision. Pour apporter une solution à ces problèmes liés à la qualité de l'information, nous proposons deux approches de fusion basées sur la théorie des croyances de Dempster-Shafer. La première approche est une mesure permettant de quantifier l'incertitude dans l'information. Cette mesure est basée sur l'entropie des croyances, et tient compte des insuffisances des mesures existantes. La deuxième approche dénommée ABC_IoT, est une méthode de pré-traitement pour étudier le problème de la véracité des informations dans les applications IoT. Le principe de l'ABC_IoT, pour vérifier la véracité des informations multisources, consiste à utiliser des informations issues de plusieurs sources pour améliorer la prise de décision. Pour ce faire, nous utilisons deux éléments nécessaires pour une bonne qualité des informations : la quantification des incertitudes et la fiabilité des sources d'information. Concernant la quantification de l'incertitude, nous utilisons la nouvelle mesure de l'incertitude. De plus, nous utilisons deux notions de fiabilité pour évaluer la confiance au niveau des sources d'information. Nous illustrons nos contributions dans le domaine des villes intelligentes où l'utilisation de l'ABC_IoT permet d'assurer une meilleure prise de décision.

Mots clés : *Internet des Objets, Qualité de l'information, Fusion d'information, Théorie de Dempster-Shafer, Entropie des croyances, Incertitude, Fiabilité, Véracité*

Remerciements

Avant toute chose, je rends grâce à *DIEU TOUT PUISSANT*, qui m'a fait la grâce de réaliser cette thèse.

Mes remerciements au Professeur *OUMTANAGA Souleymane*, mon Directeur de thèse qui m'a accueilli depuis le MASTER au LARIT, au sein duquel j'ai effectué mes premiers pas dans la recherche.

J'exprime mes sincères remerciements et toute ma gratitude au Professeur *BABRI Michel*, Directeur du LARIT, pour ses précieux conseils et sa grande générosité.

Mes remerciements vont à l'égard de tous les membres du Conseil Scientifique du LARIT dirigé par le Professeur *LOUM Georges*. Merci infiniment pour votre contribution à l'encadrement des stagiaires et doctorants du LARIT.

Merci en particulier à *FABRE Marie-Noëlle* pour son soutien indéfectible, sa confiance et ses prières qui m'ont servi de motivation pour surmonter toutes les difficultés liées à cette thèse. Toute ma gratitude au grand-frère *OUATTARA Nouho* de m'avoir fait connaître le LARIT. Merci à tous les frères doctorants, postulants et stagiaires du LARIT pour l'esprit de fraternité qui règne dans ce Laboratoire.

Enfin, j'adresse mes sincères remerciements à mes familles biologique et spirituelle pour leurs soutiens sans faille et à tous ceux et celles, qui de près ou de loin, ont contribué à la réalisation de cette thèse.

Table des matières

Résumé II

Abstract III

Remerciements IV

Table des figures IX

Liste des tableaux X

Liste des abbréviations XI

Introduction générale 1

Contexte et motivation 1

Contributions de la thèse 3

Plan de la thèse 4

1 Généralités sur l'Internet des Objets et la qualité de l'information 5

1.1 Introduction 5

1.2 Concept de l'IoT 6

1.2.1 Historique 6

1.2.2 Définition . . . 8
1.2.3 Quelques prévisions et estimations de l'IoT . . . 8
1.2.4 Architecture de l'IoT . . . 10
1.2.5 Domaines d'application de l'Internet des Objets . . . 13
1.2.6 Principaux défis dans le concept de l'IoT . . . 17
1.3 Qualité de l'information . . . 18
1.4 Fusion d'informations . . . 19
1.4.1 Définitions . . . 19
1.4.2 Architecture de la fusion d'informations . . . 20
1.4.3 Types et origines de l'information . . . 21
1.4.4 Objectifs de la fusion . . . 23
1.4.5 Représentations succinctes et explicites . . . 27
1.4.6 Imperfections de l'information . . . 29
1.4.7 Méthodes de fusion d'informations . . . 31
1.5 Conclusion . . . 33

2 La Théorie de Dempster-Shafer 34
2.1 Introduction . . . 34
2.2 Concepts de base . . . 35
2.2.1 Quelques notations . . . 35
2.2.2 Élément de preuve . . . 35
2.2.3 Croyance . . . 37
2.2.4 Plausibilité . . . 38
2.2.5 Règles de combinaison . . . 39
2.2.6 Décision . . . 41
2.3 Défis de la théorie de Dempster-Shafer . . . 42
2.3.1 La quantification de l'incertitude . . . 43
2.3.2 Gestion des conflits . . . 43
2.4 Conclusion . . . 43

3 Incertitudes et conflits de sources d'information dans les systèmes distribués 44
3.1 Introduction . . . 44
3.2 Mesures d'incertitude . . . 45

3.2.1 Quelques mesures d'incertitude dans la théorie de l'évidence de Dempster-Shafer . . . 45
3.2.2 Autres modèles . . . 47
3.3 Méthodes de résolution de conflits . . . 50
3.3.1 Méthodes basées sur la théorie des croyances . . . 50
3.3.2 Méthodes de découverte de vérité . . . 51
3.4 Conclusion . . . 52

4 Proposition d'une nouvelle mesure de l'incertitude 54
4.1 Introduction . . . 54
4.2 Analyse et formulation du problème . . . 56
4.2.1 Limites de l'entropie de Deng . . . 56
4.2.2 Limite de l'entropie proposée par Zhou *et al.* . . . 57
4.2.3 Analyse des résultats . . . 58
4.3 Nouvelle mesure de l'incertitude . . . 59
4.4 Exemples numériques et discussions . . . 61
4.4.1 Concordance avec différentes mesures d'incertitude . . . 62
4.4.2 Supériorité de la nouvelle mesure de l'incertitude . . . 63
4.5 Conclusion . . . 66

5 Une approche basée sur l'entropie des croyances pour la résolution de conflits en environnement IoT 67
5.1 Introduction . . . 68
5.2 Motivations et formulation du problème . . . 69
5.3 Notre approche de résolution de conflit (ABC_IoT) . . . 70
5.3.1 Initialisation . . . 72
5.3.2 Quantification de l'incertitude . . . 72
5.3.3 Estimation de la fiabilité . . . 73
5.3.4 Condition d'évidence . . . 74
5.3.5 Affaiblissement initial . . . 74
5.3.6 Affaiblissement final . . . 75
5.3.7 Prise de décision . . . 75
5.4 Expérimentations . . . 75
5.4.1 Cas d'un système intelligent de détection de cibles . . . 75

5.4.2 Cas d'un système intelligent dans le domaine des Villes Intelligentes 82
5.5 Conclusion 85

6 Application **86**
6.1 Introduction 86
6.2 Présentation de OpenIoT 87
6.3 Mise en œuvre et résultats 88
6.3.1 Mise en œuvre 88
6.4 Conclusion 95

Conclusion générale **97**

Liste des publications **100**

Bibliographie **109**

Table des figures

1 Illustration de la problématique 2

1.1 Estimation du nombre d'objets connectés par rapport à la population mondiale 9
1.2 Estimations du chiffre d'affaire de l'IoT 9
1.3 Architecture de l'IoT 10
1.4 Représentation du bloc FUSION 21

3.1 Méthode ECRM 50

4.1 Comparaison entre l'entropie de Zhou *et al.* et l'entropie modifiée 65
4.2 Comparaison entre l'entropie de Deng et la nouvelle entropie 65
4.3 Comparaison de la nouvelle mesure avec d'autres mesures d'incertitude 66

5.1 Logigramme de l'ABC_IoT 71
5.2 Résultat de $m(A)$ basé sur différentes règles de combinaison. 80
5.3 Résultat de $m(B)$ basé sur différentes règles de combinaison. 81
5.4 Résultat de $m(C)$ basé sur différentes règles de combinaison. 81
5.5 Résultat de $m(AC)$ basé sur différentes règles de combinaison. 82
5.6 Précision de l'ABC_IoT 84

6.1 Architecture de OpenIoT 87
6.2 Durée de recherche et d'extraction des informations issues des sources . 94
6.3 Durée de traitement 94
6.4 Résultat pour la prise de décision 95

Liste des tableaux

1.1 Evolution de l'IoT de 1999 à 2018 7
1.2 Analyse comparative des techniques de fusion 32

2.1 Les degrés de croyance calculés à partir de la distribution de masses . . 37
2.2 Les degrés de plausibilité calculées à partir de la distribution de masses 39

4.1 Mesures d'incertitude calculées à partir de l'exemple 4.1 60
4.2 Efficacité de la nouvelle mesure 61
4.3 Résultats des différentes mesures d'incertitudes des $(BOEs)$ m_3 et m_4 . . 61
4.4 Mesures d'entropie avec un sous-ensemble E d'éléments variables 64

5.1 Comparaison avec des méthodes utilisant la théorie de Dempster-Shafer 79
5.2 Différentes valeurs de précision 84

6.1 Mesures des capteurs 89
6.2 Informations issues des sources (1) 91
6.3 Informations issues des sources (2) 92
6.4 Informations issues des sources (3) 93
6.5 Mesures de température retenues après fusion 95

Liste des abbréviations

LARIT : Laboratoire de Recherche en Informatique et Télécommunication
UIT : Union Internationale des Télécommunications
FAO : Food and Agriculture Organization of the United Nations
IoT : Internet of Things
IIoT : Industrial Internet of Things
QoI : Quality Of Information
IPv6 : Internet Protocol version 6
TCP/IP : Transmission Control Protocol / Internet Protocol
DS : Demspter-Shafer
TDS : Théorie de Dempster-Shafer
FOD : Frame Of Discernment
BOE : Body Of Element
BPA : Basic Probability Assignment
BDE : Base de Données Evidentielle
ECRM : Evidential Conflict Resolution Method
ABC_IoT : Approach Based Conflict resolution in *IoT* environment
IoTSF : Internet of Things Security Foundation
MQTT : Message Queuing Telemetry Transport
IBM : International Business Machines Corporation
GSI : Global Standards Initiative
UMTS : Universal Mobile Telecommunications Service
LTE : Long Term Evolution
LAN : Local Area Network
WAN : Wide Area Network

Introduction générale

Contexte et motivation

L'émergence des technologies de l'information et de la communication (*TICs*) a favorisé l'apparition de plusieurs concepts. L'Internet des Objets *en anglais, Internet of Things - IoT*, est l'un de ces concepts. Il constitue un environnement où la collecte et le traitement des informations issues de plusieurs sources d'information par l'intermédiaire de capteurs permettent de gérer efficacement les ressources en réalisant des actions autonomes. Selon les prévisions de Cisco [1], plus de 50 milliards de dispositifs seront connectés en 2025. L'*IoT* est classé aujourd'hui parmi les plus importants producteurs des BigData.

La qualité de l'information (*QoI*) est un élément fondamental de la gestion des échanges d'informations dans L'Internet des Objets [2]. Elle repose sur des caractéristiques certaines, fiables et pertinentes mais aussi, sur l'ensemble des processus qui permettent de garantir ces caractéristiques [3]. Dans l'*IoT*, les échanges portent principalement sur la recherche d'informations auprès d'une ou plusieurs sources ou sur le fait d'agir en tant que source et de fournir des informations à une partie qui demande l'information. Cependant, le besoin de traiter et de raisonner avec les informations est nécessaire car l'information est souvent entachée de diverses imperfections [4] pouvant impacter négativement sur la performance des services et sur la prise de décision.

La fusion de diverses informations pour améliorer la qualité de l'information et la prise de décision revêtirait une importance primordiale [5–7]. La fusion d'information

est définie comme "la théorie, les techniques et les outils utilisés pour combiner des informations issues de capteurs, ou des données dérivées de données sensorielles, dans un format de représentation commun [8]".

Dans cette thèse, nous nous intéressons aux problèmes liés à l'incertitude et à la véracité des informations dans la fusion d'informations multi-sources en environnement IoT comme présenté à la figure 1. Cette figure présente un système IoT connecté à plusieurs objets ou sources d'information. Les problèmes liés à la qualité de l'information pouvant impacter négativement la prise de décision, il s'agit de répondre à la question suivante : quelles stratégies mettre en place pour une meilleure gestion des informations en environnement IoT?

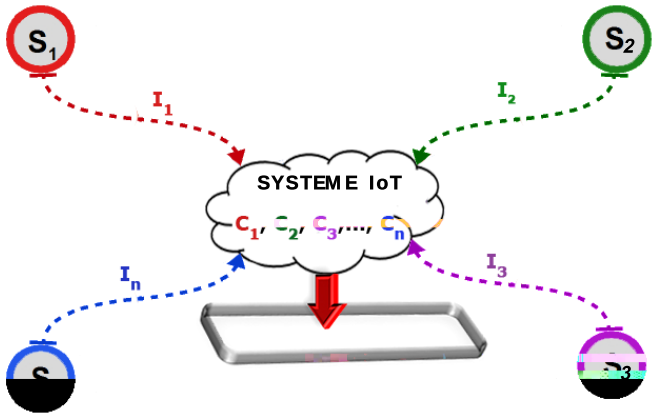

FIGURE 1 – Illustration de la problématique

Formellement, ce problème se présente comme suit. On considère n sources d'informations $(S_1, S_2, ..., S_n)$. Ces sources peuvent fournir respectivement, des informations $(I_1, I_2, ..., I_n)$ homogènes et/ou hétérogènes. Le système, chargé de la gestion de ces différentes informations, doit être capable de tirer respectivement un ensemble de conclusions $(C_1, C_2, ..., C_n)$ dont la fusion permet d'aboutir à une prise de décision P. Une conclusion peut représenter par exemple le degré de croyance d'une information. De ce fait, cela nous amène à répondre aux questions suivantes :

— Comment quantifier le degré d'incertitude lorsqu'une information est disponible?

— Comment aboutir à une prise de décision fiable et efficace lorsque l'on dispose

de plusieurs sources d'information?

Contributions de la thèse

Pour apporter une solution à ces problèmes liés à la qualité de l'information en environnement IoT, une approche fondamentale est nécessaire. Ainsi, modéliser l'incertitude dans les informations et la fiabilité des sources d'information pourrait permettre une prise décision efficace pour mieux gérer les problèmes liés à la QoI. A cet effet, nous présentons les principales contributions réalisées dans cette thèse pour tenter de répondre à cette problématique liée à la gestion de l'information en environnement IoT. En effet, modéliser la qualité de l'information revient à modéliser trois éléments : l'incertitude, la fiabilité et la véracité. Parmi les techniques de fusion d'information, la théorie de Dempster-Shafer, appelée aussi théorie des croyances ou encore théorie de l'évidence a été l'outil fondamental qui nous a permis de modéliser ces trois (03) éléments, lesquelles ont abouti à deux (02) contributions. Cette section présente brièvement ces contributions.

Proposition d'une nouvelle mesure de l'incertitude

La modélisation de l'incertitude à travers les techniques de fusion d'informations est une question importante dans les applications IoT. En effet, plusieurs facteurs liés à l'incertitude (données bruitées, erronées, manquantes, ...) peuvent entraver la prise de décision. Mesurer le degré d'incertitude des données provenant des sources d'informations avant d'appliquer la fusion de données constitue un réel défi. Pour quantifier l'incertitude dans le modèle de Dempster-Shafer, plusieurs mesures ont été proposées dans la littérature. Ces mesures parfois sous formes étendues ou hybrides utilisent le principe de l'entropie de Shannon pour la mesure de l'incertitude. Nous avons proposé une nouvelle mesure de l'incertitude qui tient compte des insuffisances des mesures existantes dans le modèle de Dempster-Shafer.

Proposition de ABC_IoT, une approche basée sur la résolution de conflits en environnement IoT

Dans cette contribution, nous abordons le problème de la véracité des informations. En environnement IoT, une technique pour s'assurer de la véracité des informations multisources, consiste à résoudre le problème de conflits lorsque ces sources transmettent des informations contradictoires. Pour améliorer la prise de décision en environnement IoT, nous avons proposé une approche basée sur l'entropie des croyances pour la résolution de conflits dans l'Internet des objets (ABC_IoT). Cette approche utilise l'incertitude des informations et la fiabilité des sources d'informations pour vérifier la véracité des informations dans l'IoT. Après analyse des différentes méthodes existantes, nous avons montré la performance et l'efficacité de l'ABC_IoT.

Plan de la thèse

La suite de ce manuscrit est organisée comme suit :

1. Le chapitre 1 présente quelques généralités sur l'Internet des Objets, la qualité de l'information et la fusion d'informations.
2. Le chapitre 2 présente la théorie de Dempster-Shafer.
3. Le chapitre 3 est un état de l'art sur les techniques pour gérer les incertitudes et conflits de sources d'information dans les systèmes distribués.
4. Le chapitre 4 présente notre première contribution, la proposition d'une nouvelle mesure de quantification de l'incertitude.
5. Le chapitre 5 présente notre deuxième contribution, une approche pour quantifier la véracité des informations issues de plusieurs sources d'information.
6. Dans le chapitre 6 nous illustrons nos contributions à travers une application dans le domaine des villes intelligentes.

Une Conclusion générale présente le bilan et quelques perspectives à ce travail.

1

Généralités sur l'Internet des Objets et la qualité de l'information

1.1 Introduction 5
1.2 Concept de l'IoT 6
1.3 Qualité de l'information 18
1.4 Fusion d'informations 19
1.5 Conclusion 33

1.1 Introduction

L'Internet des Objets en anglais, "Internet of Things, *IoT*" crée un monde dans lequel les objets physiques peuvent communiquer entre eux ou avec les humains de manière intelligente. Ces objets peuvent réagir aux changements qui surviennent dans

l'environnement et devenir des participants actifs dans ce monde intelligent. Ce chapitre aborde les généralités du concept de l'Internet des Objets et les notions qui découlent de la qualité de l'information. Ce chapitre est organisé comme suit. La section 1.2 présente le concept de l'Internet des Objets. La section 1.3 présente des notions liées à la qualité de l'information. La section 1.3 donne un aperçu sur la fusion d'informations. La section 1.5 conclut ce chapitre.

1.2 Concept de l'IoT

1.2.1 Historique

L'émergence de l'IoT est l'un des phénomènes remarquables de l'histoire de l'informatique [9]. Le terme IoT a été inventé en 1999 par Kevin Ashton [10], l'un des co-fondateurs du centre de l'identification automatique sis à l'Institut de Technologie de l'Etat de Massachusetts aux États-Unis. Il a décrit l'IoT comme un système d'interconnexion entre le monde physique et l'Internet grâce à l'utilisation de la RFID et de capteurs omniprésents qui observent et identifient le monde réel. Mais, le concept d'interconnectivité entre les dispositifs intelligents est apparu au début des années 1980 lorsqu'une machine à boisson, à l'Université de Carnegie Mellon, aux États-Unis, a été connectée à Internet pour vérifier et déclarer l'inventaire de la disponibilité des boissons [1].

En 1991, la vision actuelle de l'IoT a été donnée par Mark Weiser dans son article sur l'informatique omniprésente - The Computer of the 21st Century, traduit en français par, l'ordinateur du XXIe siècle [11]. Le tableau 1.1 montre les principaux événements associés à la croissance de l'IoT entre 1999 et 2018.

Aujourd'hui, l'IoT offre un ensemble d'opportunités à tous ses acteurs (c'est à dire, les fabricants de matériel informatique, les développeurs d'applications, les fournisseurs de services Internet, les chercheurs et industriels). Ces opportunités sont associées à l'évolution de l'Internet traditionnel, qui fournit des services et des applications pour le transfert d'informations, la communication et l'analyse dans un environnement informatique omniprésent dans lequel, les objets physiques peuvent être contrôlés et

l'environnement peut être détecté. En 2005, l'UIT a proposé que l'*IoT* établisse une connectivité intelligente et des capacités de détection parmi les objets du monde réel. Depuis lors, l'utilisation de l'*IoT* s'est développée et les grandes entreprises ont prédit une augmentation de la fabrication d'objets dotés de capteurs [1,2].

TABLE 1.1 – Evolution de l'IoT de 1999 à 2018 [11]

Année	**Evénement**
1999	Invention du terme "Internet des objets". Le premier protocole de la communication M2M dénommé MQTT [1] a été inventé.
2000	LG présente le premier réfrigérateur connecté à Internet.
2001	La Fondation Nationale de la Science aux États-Unis crée des techniques d'analyses prédictives basée sur l'*IoT*.
2005	L'Union internationale des télécommunications (UIT) [2] publie son premier rapport sur l'*IoT*.
2008	Le nombre d'appareils connectés à Internet est supérieur au nombre de personnes dans le monde.
2011	Création du groupe de réflexion, IoT Global Standards Initiative (GSI), pour la normalisation de l'*IoT*.
2012	Lancement mondial d'IPv6.
2015	Lancement de la fondation IoTSF qui lutte pour la sécurité dans l'*IoT*.
2016	Un malware dénommé "Mirai" a été utilisé pour mener une attaque DDoS par des dispositifs *IoT*.
2018	*IoTOne* [3] a créé une base de données des termes de l'*IoT*.

MQTT (Message Queuing Telemetry Transport) est un protocole de messagerie publish-subscribe basé sur le protocole TCP/IP et développé par Andy Stanford-Clark (IBM) et Arlen Nipper (EuroTech).Il est utilisé pour les appareils à ressources limitées et les réseaux à faible bande passante

L'UIT est une agence des Nations unies pour le développement spécialisé dans les technologies de l'information et de la communication, basée à Genève en Suisse

IoTOne est une source d'informations structurées sur l'Internet des objets dans le domaine de l'industrie (IIoT)

1.2.2 Définition

Plusieurs définitions ont été abordées dans la littérature. Raun a décrit l'*IoT* comme : " Un monde où le réel, le numérique et le virtuel convergent pour créer un environnement intelligent qui rend l'énergie, les transports, la maison, les industries, les villes et de nombreux autres domaines plus intelligents " [12].

Pour le CERP-IoT, l'IoT est "Un cluster des projets européens de recherche sur l'Internet des objets" l'Internet des objets est : « une infrastructure dynamique d'un réseau global. Ce réseau global a des capacités d'auto-configuration basée sur des standards et des protocoles de communication interopérables. Dans ce réseau, les objets physiques et virtuels ont des identités, des attributs physiques, des personnalités virtuelles et des interfaces intelligentes, et ils sont intégrés au réseau d'une façon transparente ». Nous retenons la définition de l'UIT qui est la suivante [13] : "L'Internet des Objets est une infrastructure mondiale pour la société de l'information, qui permet de disposer de services évolués en interconnectant des objets (physiques ou virtuels) grâce aux technologies de l'information et de la communication interopérables existantes ou en évolution." Ainsi, le concept de l'Internet des objets vise à intégrer le monde virtuel de l'Internet avec le monde physique des objets intelligents communicants, tels que les capteurs et les actionneurs, afin d'assurer une meilleure accessibilité et exploitation des ressources du monde réel.
Selon [12], un objet à trois éléments clés :

— les données produites ou reçues, stockées ou transmises;

— les algorithmes pour traiter et réutiliser ces données;

— l'écosystème dans lequel il va réagir et s'intégrer.

1.2.3 Quelques prévisions et estimations de l'IoT

Prévoir l'avenir n'est pas une tâche facile. Différentes estimations et prévisions de l'*IoT* ont été émises par différentes entreprises. La prévision la plus populaire de la littérature est celle de Cisco estimant à plus de 50 milliards de dispositifs connectés en 2020 [12], comme le montre la figure 1.1. Sur cette dernière, on peut observer que le nombre d'appareils connectés à Internet a dépassé la population mondiale en 2008.

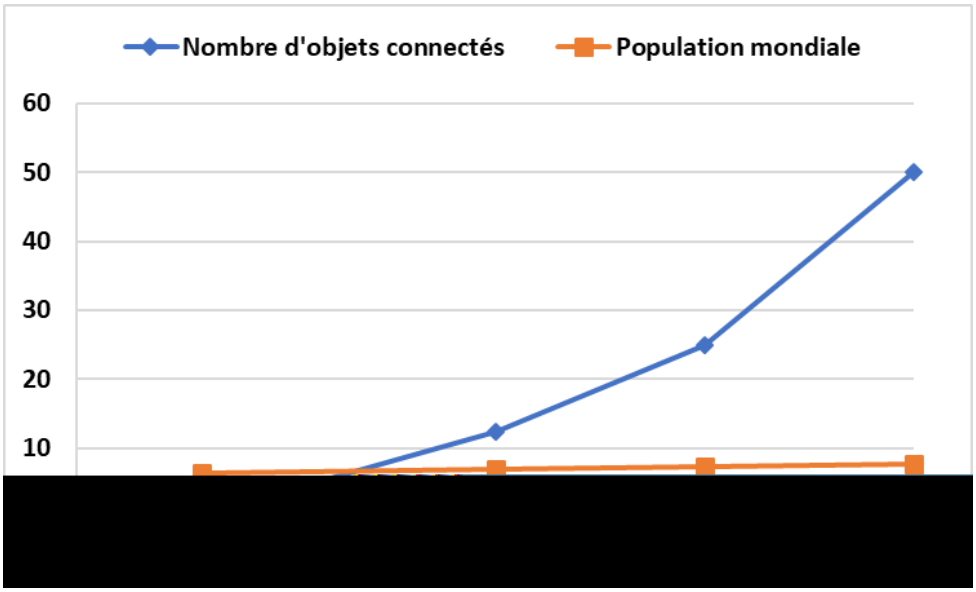

FIGURE 1.1 – Estimation du nombre d'objets connectés par rapport à la population mondiale [9]

Selon IHS [4], le nombre d'objets connectés passera de 17 milliards en 2016 à 125 milliards en 2030 [14]. Selon les données prévisionnelles de MachNation [15], le chiffre d'affaires des plate-formes *IoT* augmentera de 89% en 2019. MachNation prévoit le chiffre d'affaires de la plate-forme *IoT* à 64,6 milliards de dollars de 2017 à 2026 (Figure 1.2).

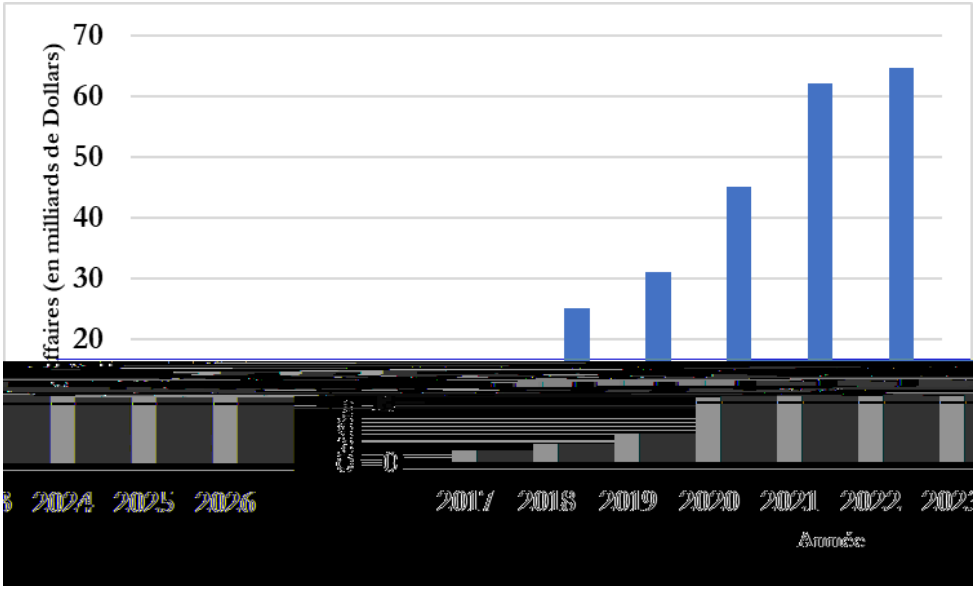

FIGURE 1.2 – [15]

4. IHS Markit est une entreprise américaine d'information économique.

1.2.4 Architecture de l'IoT

La réalisation de l'IoT est possible grâce à l'intégration de nombreuses technologies notamment les réseaux de capteurs sans-fil, les puces RFID, les réseaux personnels de faible puissance, etc. Pour comprendre le rôle clé qu'ils jouent dans l'IoT, un certain nombre de modèles pour l'IoT ont été proposés. Dans notre travail, nous nous sommes concentrés sur l'architecture proposée par l'IUT (Figure 1.3).

<table>
<tr><th>Couche</th><th>Sous-couche</th><th>Eléments de base</th><th>Technologie de base</th></tr>
<tr><td rowspan="2">Application</td><td>Applications IoT</td><td rowspan="2">Appareil portatifs, Interface & terminal utilisateur</td><td rowspan="2">Cloud Computing, Middleware, M2M, Plate-forme de service</td></tr>
<tr><td>Service de l'application</td></tr>
<tr><td rowspan="3">Transmission</td><td>LAN & WAN</td><td rowspan="3">Etablissement des connections et transmission des information</td><td rowspan="3">Internet, GPRS, WIFI, Réseau Adhoc</td></tr>
<tr><td>Réseau cœur</td></tr>
<tr><td>Réseau d'accès</td></tr>
<tr><td rowspan="2">Perception</td><td>Réseau de perception</td><td rowspan="2">Détection, identification, déclenchement & technologies de communication</td><td rowspan="2">RFIDs, WSN, GPS, Bluetooth</td></tr>
<tr><td>Nœuds de perception</td></tr>
<tr><td>Gestion du réseau</td><td colspan="2">Gestion de la sécurité des dispositifs</td><td>Gestion de la confiance</td></tr>
</table>

FIGURE 1.3 – Architecture de l'IoT [9]

Selon cette architecture, l'IoT peut être divisé en trois (03) couches : la couche de perception, la couche de transmission et la couche d'application. Ces couches ont été divisées en sous-couches selon leur fonctionnalité. Ces questions sont abordées dans les sous-sections suivantes.

La couche de perception

Elle peut aussi être appelée "couche de dispositifs", "couche sensorielle" ou "couche de reconnaissance". Cette couche est considérée comme la plus basse de l'architecture de l'IoT. Elle comprend les technologies utilisées pour la détection (collecte des données de l'environnement et leur envoi dans des bases de données, des entrepôts de données ou des nuages), l'identification (identification des objets sur la base de l'identité unique qui leur est attribuée), l'actionnement (exécution d'une action mécanique

basée sur les données détectées) et la communication (établissement de la connectivité entre dispositifs intelligents hétérogènes). Elle se caractérise par la capture de l'information du monde réel et sa représentation sous forme numérique. Selon la fonctionnalité qu'elle dessert, cette couche peut être divisée en deux sous-couches - Noeuds de perception (ou noeuds sensoriels) et Réseau de perception.

— Nœuds de perception : Il s'agit notamment des dispositifs ou objets physiques tels que capteurs, actionneurs, contrôleurs, etc. Ces dispositifs peuvent former un réseau adhoc, un réseau maillé ou un environnement à sauts multiples afin d'augmenter l'évolutivité et d'accélérer le déploiement;

— Réseau de perception : Il est responsable de la communication avec le réseau de transport. Il transmet de manière sécurisée les données collectées par les nœuds de perception à des passerelles pour une transmission ultérieure et envoie des signaux de commande aux dispositifs de commande par l'intermédiaire d'un moyen de communication filaire ou sans fil.

La couche de transmission

La couche de transmission peut également être appelée "couche de transport" ou "couche réseau". Elle est l'interface entre la couche perception et la couche application. Elle peut être considérée comme l'intégration d'une variété de réseaux, de technologies et de protocoles hétérogènes. Son but est de transmettre les données recueillies par les nœuds de perception vers les unités de traitement de l'information (ou les unités décisionnelles de haut niveau) au moyen de canaux de communication avec ou sans fil pour l'analyse, l'exploration, l'agrégation et le codage des données [3]. Selon la fonctionnalité, elle peut être divisée en trois sous-couches : le réseau d'accès, le réseau central et le réseau local et étendu.

— Le réseau d'accès : Il fournit un environnement d'accès ubiquitaire pour la couche de perception [16]. Il s'agit d'un type de réseau de télécommunication qui sert de pont entre l'abonné et les fournisseurs de services. Il établit des capacités de communication et d'infrastructure comme la communication mobile, la communication par satellite et la communication sans fil pour les auditeurs. Les réseaux d'accès que l'IoT peut mettre en œuvre sont le réseau adhoc, le réseau GPRS, le réseau 2G, 3G, 4G-LTE, le réseau WiFi, ZigBee, LowEnergy Bluetooth, etc;

— Le réseau central : Le réseau central est l'Internet, qui fournit le cadre de base de l'IoT. Il est responsable de la transmission des données aux utilisateurs finaux qui sont connectés par le réseau d'accès. C'est la partie centrale de tout réseau de télécommunication qui sert d'épine dorsale pour l'échange d'informations et de services. Il établit la communication entre les dispositifs restreints pour le partage des ressources. Internet peut être utilisé comme réseau public ou privé, réseau d'affaires ou réseau gouvernemental et a une portée locale et étendue. Il permet d'observer et de contrôler les objets physiques à distance;

— Réseau local et étendu : Le réseau local (LAN) est une interconnexion entre des dispositifs à l'intérieur d'une région relativement plus petite. Les appareils d'un réseau local peuvent communiquer directement entre eux et des appareils distants à l'aide de passerelles. De la même manière, les réseaux étendus (WAN) sont considérés comme une distribution de dispositifs sur des zones géographiques plus vastes.

La couche Application

C'est la couche supérieure de l'architecture de l'IoT, visible par l'utilisateur final. Le but de cette couche applicative est de gérer et de fournir les applications sur la base des informations collectées par la couche de perception et traitées par l'unité de traitement de l'information. Elle permet aux utilisateurs finaux d'accéder à des services personnalisés sur le réseau en fonction de leurs besoins grâce à l'utilisation de divers appareils portatifs et équipements terminaux. Il peut être divisé en deux sous-couches : la couche de support aux applications et les applications IoT.

— La couche de support aux applications : Elle est juste au-dessus de la couche Transmission. Elle fournit un soutien à une variété de services et est responsable de l'exécution de calculs intelligents et du traitement ou la fusion des données. Elle effectue la reconnaissance et la filtration des données afin de les catégoriser en données valides, invalides, malveillantes, spam, etc. Selon les différents services, cette couche peut être organisée de différentes manières. Elle exploite les fonctionnalités du middleware, qui est responsable des calculs intelligents;

— Applications de l'IoT : les applications IoT peuvent être divisées en trois catégories sur la base de leurs fonctionnalités - la collecte d'informations, l'analyse

et les prises de décision. Les applications collectives d'information sont chargées de collecter les données des nœuds de perception et de leur stockage local, l'analyse traite du prétraitement hors ligne des données collectées pour créer un modèle générique pour l'évaluation des données qui seront collectées à l'avenir et les applications décisionnelles sont utilisées pour prendre les mesures appropriées selon les données analysées.

1.2.5 Domaines d'application de l'Internet des Objets

Le nombre des applications de l'IoT augmentent de jour en jour. Qu'il s'agisse d'une petite pilule ou d'un gros avion, de simples bouilloires intelligentes ou de véhicules complexes sans homme, tous ces objets font désormais partie intégrante de l'IoT. Les principaux domaines d'application de l'Internet des Objets sont la domotique, la santé, le transport, l'agriculture, les Villes Intelligentes [2].

La domotique

La domotique ou "maison connectée" est vue comme l'apogée de l'Internet des objets [17]. Ce concept porte le nom de "la science de la maison", qui peut être définie comme l'ensemble des appareils connectés utilisés pour l'élaboration d'un bâtiment intelligent. De plus, grâce à des plateformes dédiées nommées box domotiques, ces dispositifs peuvent interagir entre eux et voir leur contrôle centralisé sur une application unique. Ainsi, c'est toute la maison qui se retrouve connectée et contrôlable depuis son smartphone, du contrôle des lumières à la gestion de l'énergie.
De manière générale, trois utilités aux technologies domotiques sont dénombrées :

— le confort (en améliorant la qualité de vie et en automatisant certaines tâches),

— les économies d'énergie (en installant des modules qui surveilleront les dépenses des utilisateurs et leur donneront des conseils personnalisés),

— la sécurité (grâce à des systèmes d'alarme centralisés et un système de scénarios).

La domotique couvre de nombreux secteurs et permet de nombreuses fonctionnalités, en voici les principales : la création d'automatismes, le pilotage à distance, le contrôle de l'éclairage, la gestion du chauffage et la centralisation de la sécurité.

La Santé Connectée

la "Santé Connectée" est un sous-domaine de l'e-Santé qui est apparu suite au développement de l'Internet des Objets. Notons que l'e-santé est la télémédecine effectuée par le biais d'Internet. En d'autres termes, "la Santé Connectée" est l'utilisation des nouveaux moyens induits par l'Internet des Objets dans le secteur de l'e-Santé, et permettant dès lors d'accroître ses possibilités/applications. Nombreux sont les bénéfices que les technologies de l'IoT pourraient apporter au domaine de la e-Santé, et donc au domaine de la "Santé Connectée". Leurs applications résultantes peuvent être groupées en quatre ensembles généraux de bénéfices : le suivi d'objets et de personnes, l'identification et l'authentification de personnes, la collection de données.

Les transports

L'Internet des Objets a fait sentir sa présence dans presque toutes les industries et les transports ne font pas exception. L'IoT dans les transports peut aider à suivre les véhicules et les personnes, relier les infrastructures et améliorer le transport. Internet, ainsi que les applications mobiles, la communication et la technologie accrues telles que la collecte de données, la triangulation du téléphone cellulaire, la communication machine à machine, le Cloud Computing, etc., le potentiel de révolution du transport est élevé.

L'agriculture

Selon la FAO, pour nourrir les 2,3 milliards d'êtres humains supplémentaires qui peupleront la planète en 2050, nous devrons augmenter de 70% la production alimentaire. La nécessité de produire plus et plus rapidement s'impose donc comme une priorité [18]. L'Internet des Objets peut faire la différence.
En combinant IoT et l'analyse des Big Data, les exploitants peuvent obtenir une grande palette de données et s'en servir pour gagner en efficacité, accroître leur productivité et surveiller la qualité des produits alimentaires, du champ à l'assiette.
Les professionnels du secteur agricole doivent gérer et surveiller leurs productions et leurs équipements répartis sur des surfaces plus ou moins grandes : récoltes, tracteurs ou encore matériel d'irrigation. Les solutions IoT peuvent aider grandement les systèmes d'irrigation ou le matériel agricole en effectuant l'analyse des données col-

lectées par les capteurs IoT. Cela permet d'avoir une vue globale des performances et permet de mieux programmer les opérations d'entretien pour éviter tout risque de panne qui pénaliserait le rendement. Dans le domaine de l'agriculture de précision, l'accès en temps réel aux données des relevés météo, de la qualité de l'air, des sols et des niveaux d'hydratation peut aider à prendre de meilleures décisions au moment des récoltes.

Les environnements intelligents

Le concept d'environnements intelligents découle de la définition de l'informatique ubiquitaire qui, selon Mark Weiser , promeut l'idée d'un "monde physique richement et invisiblement imbriqué avec des capteurs, des actionneurs, des affichages et des éléments de calcul, intégrés de manière transparente dans les objets du quotidien" de nos vies et connectés via un réseau continu ". [18]

Cook et Das [16] définissent l'environnement intelligent comme "un petit monde dans lequel différents types d' appareils intelligents travaillent en permanence pour rendre la vie des habitants plus confortable". Les environnements intelligents visent à satisfaire l'expérience des individus de tout type d'environnement, en remplaçant le travail dangereux, le travail physique et les tâches répétitives par des agents automatisés. Poslad [19] distingue trois types d'environnements intelligents pour les systèmes, les services et les périphériques: les environnements informatiques virtuels, les environnements physiques et les environnements humains :

— Les environnements informatiques virtuels ou sources d'information permettent aux appareils intelligents d'accéder aux services pertinents n'importe où et à tout moment;

— Les environnements physiques peuvent être intégrés à une variété de périphériques intelligents de différents types comprenant des capteurs et des contrôleurs;

— Environnements humains: les êtres humains, individuellement ou collectivement, forment un environnement intelligent pour les périphériques. Cependant, les humains peuvent eux-mêmes utiliser des appareils intelligents tels que les téléphones portables ou encore utiliser des appareils (ordinateurs portables) qui intègrent d'autres appareils (par exemple, des stimulateurs cardiaques pour maintenir une opération cardiaque saine ou des lentilles cornéennes).

Les environnements intelligents sont généralement classés comme présentant les caractéristiques suivantes :

— Contrôle à distance des périphériques , tels que les systèmes de communication par courant porteur pour contrôler les périphériques;

— Communication de périphérique, utilisation de middleware et communication sans fil pour former une image des environnements connectés. Acquisition / diffusion d'informations à partir de réseaux de capteurs Services améliorés par appareils intelligents;

— Capacités prédictives et décisionnelles.

En tant que nouvelle forme de développement durable, le concept de "villes intelligentes" a suscité beaucoup d'intérêt ces dernières années. Plusieurs définitions ont été proposées pour décrire ce concept. Mais, c'est encore un phénomène vague ou flou [16, 20]. Selon [4], le concept de "ville intelligente" décrit une ville moderne qui utilise une infrastructure d'information intelligente pour assurer la durabilité et la compétitivité des différentes fonctions urbaines en intégrant différentes dimensions du développement urbain et des investissements. Le premier modèle de ville intelligente a été proposé par Cisco à Dubaï [20]. Cisco permet à Dubaï un gouvernement intelligent [21]. Un autre modèle de villes intelligentes a été proposé par IBM à New York [22]. Dans ce contexte, IBM fournit un ensemble d'applications, telles que : Le réseau de gestion intelligente des transports, la gestion intelligente des bâtiments, la gestion intelligente des ressources en eau, etc. Siemens, propose également un modèle de ville intelligente en Allemagne, similaire au modèle de ville intelligente de Montréal [23]. Par conséquent, plusieurs modèles de villes intelligentes ont été proposés et tous ces modèles ont les mêmes composantes. On parle d'économie intelligente, d'environnement intelligent, de Population intelligente, etc.

Outre les domaines sus-mentionnés, l'Internet des Objets touche aussi les domaines de l'assurance, du commerce, des loisirs et les *Wearables* (vêtement ou accessoire comportant des éléments informatiques et électroniques avancés), et bien d'autres.

1.2.6 Principaux défis dans le concept de l'IoT

Beaucoup de recherches importantes, dans de nombreuses directions, sont nécessaires pour une utilisation à l'échelle de l'IoT [24]. Dans cette section, nous évoquons les principaux enjeux du concept de l'IoT.

L'adressage

La trajectoire actuelle du nombre de dispositifs intelligents déployés implique qu'à terme, des milliards d'objets se retrouveront sur Internet. Comment nommer, authentifier, maintenir et protéger les accès si un grand nombre d'objets ont des problèmes majeurs [25]. L'$IPv6$ suffira-t-il ? Les protocoles tels que $6LowPAN$ joueront-ils un rôle ? Des normes et des protocoles entièrement nouveaux verront-ils le jour ? Étant donné que beaucoup d'objets sur Internet nécessiteront leur propre source d'énergie, est-ce que les circuits de balayage d'énergie de très faible puissance élimineront le besoin de piles ? Comment les quantités massives de données seront-elles recueillies ?

La sécurité

Un problème fondamental qui est toujours présent dans l'Internet d'aujourd'hui et qui doit être résolu est celui de la sécurité [2, 26]. Les attaques sont problématiques pour l'IoT en raison de la capacité minimale des "objets" (dispositifs) utilisés, de l'accessibilité physique aux capteurs. Le problème de sécurité est de plus en plus présent par le fait que les défaillances aléatoires des objets constituent des vulnérabilités qui peuvent être exploitées par les attaquants. Pour répondre aux exigences réalistes du système qui découlent d'un fonctionnement de longue durée et sans surveillance, les applications IoT doivent pouvoir continuer à fonctionner de manière satisfaisante en présence d'attaques de sécurité et à se rétablir efficacement de celles-ci. Les solutions peuvent nécessiter le téléchargement d'un nouveau code et ceci est lui-même ouvert aux attaques de sécurité. Le système doit également être capable de s'adapter à de nouvelles attaques imprévues lors du déploiement initial du système.

La gestion des données

Dans le monde des objets connectés, il existe une grande quantité de données brutes collectées en continu. Il est nécessaire de développer des techniques qui conver-

tissent ces données brutes en connaissances utilisables. Par exemple, dans le domaine médical, les flux bruts de valeurs des capteurs doivent être convertis en activités significatives effectuées par une personne, comme manger, mal respirer ou présenter des signes de dépression. Selon [25], les principaux défis pour l'interprétation des données et la formation des connaissances comprennent l'étude des données imparfaites et physiques du monde et le développement de nouvelles techniques d'inférence qui peuvent être exploités par les réseaux bayésiens ou la théorie de Dempster-Shafer. Ces limites nécessitent le besoin de connaître les probabilités a priori et le coût des calculs. Notre thèse est principalement fondée sur ce défi présenté par [25].

1.3 Qualité de l'information

La qualité de l'information est un concept important en théorie de l'information [27, 28]. Elle constitue un sujet de recherche apparu depuis le développement des techniques modernes de traitement numérique [29]. Selon Ravi et al. [2], étudier la qualité de l'information revient à étudier les problèmes qui sont des sources de dégradation de l'information. Dans la littérature, les auteurs soulignent sept (07) facteurs spécifiques qui dégradent la *QoI* dans les systèmes distribués [2, 6, 29]. Ce sont :

1. La propagation des incertitudes : Une information peut provenir d'un certain nombre de processus. A chaque processus, une incertitude ou un bruit supplémentaire peut s'ajouter aux valeurs existantes dégradant ainsi l'information;

2. La véracité ou la crédibilité des sources d'information : Il est nécessaire de quantifier la véracité des différentes sources d'information. Par exemple, la degré de confiance d'une source qui rapporte une information est différent des autres ;

3. L'exactitude ou précision des capteurs : Les capteurs peuvent ou non fournir des informations précises pour un domaine donné en raison de diverses raisons notamment, la défectuosité des capteurs, l'influence de l'environnement sur les capteurs, etc. La précision de l'information peut ne pas être suffisante pour le domaine d'application ;

4. Le délai de diffusion de l'information : Il peut y avoir un décalage entre la survenance d'un événement et son stockage dans le système réel ;

5. L'absence de données : Cela peut provenir des capteurs qui ne répondent pas ou qui ne sont pas déployés dans la zone concernée ;

6. La discordance dans la représentation de l'information : La représentation utilisée pour stocker l'information peut ne pas convenir au domaine d'application ;

7. L'adaptation à l'usage : Les données capturées pour un usage particulier peuvent ne pas convenir à d'autres fins ;

1.4 Fusion d'informations

La fusion d'informations désigne l'ensemble des méthodes pour gérer des quantités très importantes de données multisources [8, 30, 31]. Depuis quelques années des méthodes de fusion ont été adaptées et développées pour des applications en théorie de l'information. Il s'agit ici d'améliorer la connaissance du monde observé pour le décrire du mieux possible tel qu'il est [8]. Dans cette section, nous présentons quelques notions importantes de ce concept.

1.4.1 Définitions

Compte tenu de l'éventail des problèmes d'application et des solutions proposées, il existe des définitions très diverses de la notion de fusion en informatique. Historiquement, un premier groupe de travail, le *Joint Directors of Laboratories* (JDL) du département de la défense américaine s'est créé en 1986. Ce groupe s'est intéressé à la définition de la terminologie de la fusion d'informations [8].

Définition 1.4.1

> La fusion est un processus multi-niveaux et à facettes multiples ayant pour objet la détection automatique, l'association, la corrélation, l'estimation et la combinaison d'informations de sources singulières et plurielles.

Le groupe de travail européen FUSION a défini la fusion d'informations dans un cadre conceptuel plus général [31].

Bloch [32] propose une définition plus spécifique. Nous constatons que dans toutes ces définitions, l'information est issue de plusieurs sources. Dans cette situation, nous

Définition 1.4.2

> La fusion d'informations consiste à regrouper des informations issues de plusieurs sources d'informations et à exploiter l'information regroupée.

Définition 1.4.3

> La fusion d'informations consiste à combiner des informations issues de plusieurs sources afin d'améliorer la prise de décision.

nous referons plus à la définition 1.4.3. Cette définition contient deux ingrédients principaux. Premièrement, elle met l'accent sur la combinaison de l'information. Deuxièmement, elle met l'accent sur les objectifs de la fusion.

1.4.2 Architecture de la fusion d'informations

Selon L. Walid et al. [33], Il est intéressant d'obtenir à la sortie de la fusion un indice de qualité en plus du résultat issu d'une décision qui dépend de l'application. Cet indice de qualité permettra d'évaluer l'opération de fusion et de compléter les informations supplémentaires. De plus, ils indiquent que les connaissances externes au système permettent le choix d'une modélisation adéquate pour l'opération de fusion. Ils définissent quatre étapes (Figure 1.4) dans le processus de fusion d'informations :

— Modélisation

La modélisation est la première étape ; elle consiste à choisir un formalisme (i.e. la représentation des informations à fusionner). Cette étape est donc particulièrement déterminante. La modélisation consiste à déterminer une distribution, une fonction de coût, une formule, etc.

— Estimation

L'estimation est dépendante de la modélisation, elle n'est pas une étape systématique, mais souvent nécessaire pour la plupart des formalismes. Il s'agit par exemple de l'estimation des distributions, si c'est le formalisme choisi. Pour cette étape les informations supplémentaires peuvent aider.

— Combinaison

L'étape de combinaison est la phase de regroupement des informations. C'est le cœur de la fusion. Il s'agit de choisir un opérateur adapté au formalisme de la

modélisation. Ici aussi des informations supplémentaires peuvent guider ce choix. Les opérateurs de combinaison fournissent un résultat de même nature que les informations combinées. L'information ainsi fournie s'interprète de la même façon que les informations de départ. Les propriétés intéressantes des opérateurs pour la fusion sont l'associativité, la commutativité, l'idempotence [34], ainsi que l'adaptabilité [2,30]. Une décision partielle prise avant la combinaison pourrait entraîner des contradictions. Le fait de conserver la nature de l'information permet de ne prendre aucune décision partielle avant la combinaison.

— Décision

La dernière étape de la fusion d'informations est la décision. En effet, d'après notre définition de la fusion, la combinaison est réalisée en vue d'améliorer la prise de décision. Le choix du critère se fait en fonction du formalisme (c'est à dire, du choix de la modélisation et de la combinaison).

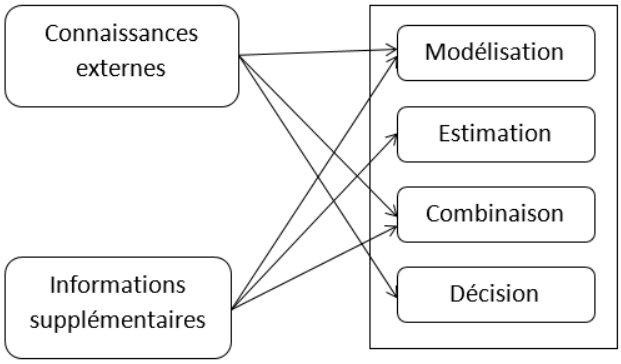

FIGURE 1.4 – Représentation du bloc FUSION [30]

Les connaissances externes et supplémentaires peuvent ici apporter une aide importante au choix du critère. Classiquement, il s'agit de la normalisation, la minimisation ou de la maximisation d'une fonction issue de la fusion. C'est cette même fonction qui peut servir au calcul d'un indice de qualité. Cette étape doit donc fournir au système la "meilleure" décision. Les manières de considérer ces quatre étapes définissent l'architecture de la fusion d'informations.

1.4.3 Types et origines de l'information

Les problèmes de fusion en général, traitent de différents types d'informations. Nous distinguons deux types d'informations. L'information descriptive qui se rap-

porte au monde réel, tandis que l'information normative se rapportant à un monde idéal. L'information descriptive comprend les observations et les connaissances. L'information normative comprend les préférences et les différents règlements.

Observations

Les types d'informations qualifiés d'observations couvrent des informations telles que des données issues de capteurs, des preuves, des faits, des mesures. En général, ce sont des éléments de description du monde (au moment où les observations sont collectées), ils se réfèrent à une situation particulière, et non à une situation générique. Ce type d'information peut être aussi bien objectif que subjectif. Le premier cas concerne des données fournies par des capteurs ou dispositifs, tandis que le second concerne des avis fournis par une personne. Selon le type d'observations, les éléments à fusionner peuvent prendre différentes formes : nombres, intervalles, distributions subjectives d'incertitude, distributions subjectives d'incertitude, certitudes qualifiées, etc.

Connaissance

La connaissance décrit comment le monde est en général. Elle ne fait pas nécessairement référence à une situation particulière en tant qu'une observation, mais plutôt à quelque chose de plus générique, comme une population ou une catégorie de situations. Les connaissances peuvent être bien définies (par exemple sous forme de données statistiques) ou non (connaissances par défaut). Les éléments à fusionner sont alors les règles (éventuellement avec des exceptions), les règles floues, les distributions de probabilités statistiques, les fonctions de vraisemblance, les ordres de plausibilité, les modèles mathématiques, les lois de la nature, etc.

Préférences

Contrairement au type d'information précédent, les préférences sont toujours des descriptions subjectives des désirs d'un individu ou d'un groupe de personnes. Ils se réfèrent à la façon dont ils aimeraient que le système soit. Les éléments à fusionner sont alors des fonctions utilitaires ou des relations de préférences binaires ou floues. Comme pour les autres types d'information, l'information sur les préférences peut être

incomplète, imprécise ou incertaine.

Réglementation

La réglementation se réfère à la façon dont le monde devrait être en général. Le monde idéal en ce sens, est celui où la réglementation est satisfaite. Ces réglementations peuvent décrire ce qui est obligatoire et ce qui est interdit dans un certain domaine d'intérêt. Il s'agit encore une fois d'un type générique d'information, qui fait référence à une classe de situations rencontrées dans ce domaine.

1.4.4 Objectifs de la fusion

Nous trouvons quatre raisons principales de combiner des informations dans la littérature [30]:

— Affiner ou élargir les connaissances, les informations ou les croyances sur le monde réel;

— Construire un point de vue global sur un monde idéal;

— Mettre à jour ou rafraîchir des informations sur le monde réel;

— Aider à la prise de décision.

Ces raisons correspondent, respectivement, à la formation ou à l'actualisation d'un modèle du monde réel ou idéal.

Former un modèle du monde réel

Une première raison de combiner est d'affiner l'information sur le monde réel. Un tel cas suppose que l'on dispose de plusieurs sources d'information, décrivant l'état actuel du monde réel, généralement en termes de valeurs d'attributs. Des exemples typiques de cette situation sont donnés par des problèmes de détection de cible. D'autres applications typiques peuvent être trouvées dans la classification multi-capteurs ou dans la reconnaissance de formes. Dans de telles applications, le but est d'étiqueter certains objets, à l'aide des informations du capteur. La fusion bénéficie de la complémentarité des capteurs impliqués, concernant leur capacité de détection. Chaque capteur peut mettre en évidence les différentes caractéristiques des objets, et un seul capteur n'est généralement pas suffisant quand on veut une classification précise et

fiable entre des objets assez similaires (c'est pourquoi la fusion est nécessaire).
D'une manière générale, la combinaison des informations fournies par les différentes sources, compte tenu de leur niveau de fiabilité respectif, permettra d'affiner les connaissances sur le monde réel : l'imprécision et l'incertitude sont réduites. Lorsqu'une combinaison conjonctive est réalisée, les informations incomplètes d'une source peuvent être complétées par d'autres sources.
Une source d'information est tout dispositif physique, ou un opérateur humain, qui fournit des données de quelque nature que ce soit, concernant les attributs décrivant le monde réel. Les données fournies par les sources instancient les valeurs des attributs. Ces valeurs peuvent être imprécises (par exemple "environ 5,4", "assez grandes"), incertaines (par exemple, "il semble s'agir d'un avion de guerre"), ou incomplètes, c'est-à-dire, certains attributs ne sont pas instanciés (par exemple, "vitesse inconnue"). De plus, les sources elles-mêmes peuvent être plus ou moins fiables. Un problème souvent mentionné est que toutes les informations fournies par les sources se rapportent au même monde (réel) (ou au même objet), de sorte qu'il est naturel de les combiner, en tenant compte de leur fiabilité et de la qualité des informations. Cela permet d'obtenir une vision plus complète du monde réel. Cependant, dans des applications pratiques comme le suivi de cibles et l'évaluation de situation, c'est loin d'être le cas. La première étape, avant toute action de combinaison, est de vérifier que les informations fournies par les sources portent sur le même objet. Cette étape préliminaire (mais la plus difficile, selon [8,35], appelée association (ou association de données dans le suivi de cibles multiples) consiste à trouver l'élément la partition la plus plausible de l'ensemble des sources, divisant ainsi l'information en des objets distincts.
La plupart du temps, le problème est rendu plus difficile par le fait que le nombre a priori d'objets observés est inconnu. Dans le suivi de cibles, la fonction d'incertitude sur l'association provient de fausses alarmes aléatoires dans la détection, l'encombrement, les cibles gênantes ou d'autres contre-mesures. Les principaux modèles utilisés dans ce domaine sont soit déterministes (impliquant des tests d'hypothèses classiques), soit déterministes ou probabiliste (s'appuyant principalement sur le cadre bayésien).

Former un modèle de monde idéal

Une deuxième raison de combiner est d'avoir un point de vue global, prenant en compte différents points de vue, tels que les préférences individuelles, ou les réglementations consensuelles sur la manière dont le monde ou le système devrait être. Ce problème diffère considérablement du précédent. Dans le raffinement des connaissances, l'objectif est d'obtenir une description plus complète et synthétique du monde réel, en éliminant les redondances et les incohérences, puis en améliorant la précision. Il ne s'agit pas nécessairement ici d'obtenir une description plus précise du monde idéal, mais plutôt d'évaluer des alternatives. Dans le cas des préférences, les sources peuvent être des experts humains, des électeurs, puis on parle de décision multi-personnelles, de mise en commun des opinions, de problème de vote, etc., puis on parle d'analyse multi-critères ou de décision. Un exemple de cette dernière catégorie est le suivant : considérons une base de données décrivant des voitures et qui prend en compte plusieurs critères d'évaluation (tels que "prix", "performance", "essence", "confort"), chaque critère donne son point de vue pour l'évaluation des voitures (l'"essence", "confort"). Chaque critère donne son point de vue pour l'évaluation des voitures (le point de vue du prix, etc.), et nous devons les combiner afin d'avoir un point de vue global qui permet de prendre une décision.

Dans le cas de la réglementation, c'est-à-dire d'un ensemble de "lois" spécifiant ce qui est permis, obligatoire ou interdit de faire, comme par dans le monde actuel par exemple, "Vous ne devez pas tuer", "Vous ne devez pas manger avec vos doigts", et ainsi de suite, les sources peuvent être considérées comme différents systèmes culturels, politiques ou religieux, chacun ayant sa propre réglementation cohérente. Ayant plusieurs points de vue différents provenant de plusieurs sources, le problème est de les combiner afin d'obtenir un point de vue global et cohérent. Il arrive rarement que toutes les sources aient plus ou moins la même opinion sur le monde, de sorte que le conflit est ici la règle. Habituellement, si le conflit entre plusieurs points de vue n'est pas trop fort, on prend une sorte de point médian entre eux, tout en considérant l'importance (niveau d'expertise, pertinence, etc.) de chaque source, ainsi que les interactions possibles entre elles. Si le conflit est trop fort, les deux sources à l'origine du conflit ne doivent pas être combinées. Il se trouve que le problème est très difficile dans le cas de la réglementation. Par exemple, que doit faire un soldat chrétien si un

règlement (loi religieuse) dit que "les chrétiens ne doivent pas tuer", et un autre (loi militaire) dit que : "les soldats doivent tuer" ?
Enfin, nous voulons souligner la différence entre les procédures de combinaison entre le cas présent et le cas d'affinement et de mise à jour des connaissances. Prenons par exemple deux sources qui donnent leur point de vue sur la préférence de a par rapport à b. Une source peut dire qu'elle préfère fortement a par rapport à b, tandis que la seconde peut dire que a et b sont indifférents. Une combinaison raisonnable des deux conduit à la conclusion que globalement, a est plus ou moins préféré à b, c'est-à-dire que nous avons fait une sorte de moyenne des deux opinions. Considérons maintenant le cas du raffinement des connaissances. Supposons que deux sources fiables estiment la position d'une cible. Une source dit que l'objet est près de A, et l'autre source dit que l'objet est près de B. A et B étant des emplacements imprécis disjoints. L'application d'un type de combinaison basé sur la moyenne, dans ce cas, conduirait à la conclusion que la véritable position de l'objet se trouve entre les deux points A et B, ce qui est clairement faux, puisque les deux sources, qui sont fiables, s'accordent sur le fait que l'objet n'est certainement pas en dehors de A ni de B. Ces exemples montrent que des situations différentes doivent être traitées dans des contextes complètement différents et différentes combinaisons doivent être utilisées.

Mise à jour d'un modèle du monde réel

Une troisième raison de combiner est de mettre à jour, de réviser ou de rafraîchir l'information sur le monde réel [36, 37]. Supposons que nous avons un modèle actuel du monde réel. Le modèle actuel est donc un ensemble d'attributs instanciés, décrivant le monde réel à un moment donné. Ce modèle peut avoir été obtenu par une combinaison antérieure de sources se rapportant au monde réel considéré. Si une nouvelle information arrive, Galland et al. [38] montrent que l'on peut la combiner avec l'information existante, en tenant compte de son niveau de fiabilité. Ici, la nature de la combinaison est essentiellement asymétrique, puisque l'on ne peut attribuer le même rôle à une source apportant de l'information nouvelle et à une base de connaissances ou d'informations a priori. A ce stade, il convient d'établir une distinction entre la révision et la mise à jour.

— La révision modifie l'information préalable sur le monde à l'instant t (référentiel de situation) par de nouvelles informations sur l'instant t avec priorité à cette der-

nière (le monde est envisagé d'un point de vue statique, mais l'asymétrie concerne le flux d'information). La modification des informations préalables est minimale;

— La mise à jour modifie les informations antérieures sur le monde au temps t par de nouvelles informations sur le temps $t+1$ avec priorité à ce dernier, de manière à décrire le monde au temps $t+1$. La mise à jour est réalisée soit par évolution minimale de chacun des états a priori possibles du monde vers un état compatible avec l'information d'entrée (par exemple, les bases de données) soit par une étape de prédiction (à l'aide d'un modèle de l'évolution du monde).

Mise à jour d'un modèle du monde idéal

Les distinctions précédentes (entre la révision et la mise à jour) sont pertinentes lorsqu'il s'agit de la préférence [39]. Bien que la différence entre la révision des préférences et la préférence n'ait pas reçu beaucoup d'attention jusqu'à présent, une idée beaucoup plus classique est l'incorporation dans une "base de préférences" d'un nouvel objectif, un nouveau règlement ou d'un fait nouveau

1.4.5 Représentations succinctes et explicites

La planification du développement d'un système d'information doté de capacités de fusion requiert une attention à la structure de données utilisée pour coder aussi efficacement que possible les informations disponibles [24]. Lorsque l'ensemble des états (qu'il s'agisse du monde réel ou d'un monde idéal) a une taille raisonnable, les éléments de connaissance ou de préférence peuvent être représentés explicitement, en énumérant toutes les situations possibles, éventuellement par ordre de classement, ou bien avec un degré représentant, par exemple, la probabilité, la possibilité, la croyance, l'utilité, le coût etc. [26]. Différents langages pour la représentation de l'information existent et peuvent être répartis en deux classes [40].

Langages de représentation logique

Ces langages sont fondés sur une logique propositionnelle ou de premier ordre, consistant par exemple soit à spécifier des contraintes numériques sur le degré d'incertitude ou de préférence des formules logiques, soit à distinguer explicitement entre

faits et obligations. Les langages de représentation logique ne sont pas seulement succincts, mais ils offrent également deux autres caractéristiques intéressantes :

— ils ont une grande expressivité et permettent de coder les données au plus près de l'intuition;

— ils sont équipés de procédures de décision afin de permettre l'automatisation du processus de fusion. Dans un cadre logique, il y a deux façons différentes de représenter l'information, qui correspondent à la théorie dite du modèle et à la théorie de la preuve de la logique. Selon la théorie du modèle, l'information est représentée par un ensemble d'états. Les conséquences d'une information sont définies par les phrases qui sont vraies dans tous ses modèles, c'est-à-dire dans tous les mondes où elle est vraie. Selon la théorie de la preuve, une information est représentée par une phrase (ou un ensemble de phrases) du langage associé à la logique.

Modèles graphiques

Cette terminologie se réfère à des langages qui expriment de manière explicite des dépendances entre variables. Par exemple, les réseaux bayésiens [19](exprimant, pour chaque variable, sa table de probabilité conditionnelle selon la valeur des variables dont elle dépend) peuvent être utilisés pour représenter succinctement une distribution de probabilité commune. Un autre exemple est le langage concernant la satisfaction des contraintes où un ensemble de contraintes permet d'obtenir une représentation succincte d'un ensemble de possibilités du monde réel ou idéal.

Représentations numériques et symboliques

L'information peut être numérique ou symbolique. L'information numérique signifie que l'on dispose des données sous la forme de nombres. Ces chiffres peuvent représenter divers attributs quantitatifs, généralement des mesures physiques, des niveaux de gris, des évaluations de caractéristiques dans des images, etc. Ils peuvent être lus directement à partir des images à fusionner, ou liées au domaine ou aux connaissances contextuelles. L'information symbolique fait allusion à toutes les informations données en termes de structures, ou de chaînes de caractères comme graphiques, termes linguistiques, propositions logiques, règles, etc. Ce type d'information peut être liée aux données à combiner ou à la connaissance du domaine.

D'une manière générale, les auteurs [41–43] montrent que les données numériques peuvent être interfacées avec des données symboliques à des fins de catégorisation et de communication. De plus, dans plusieurs situations, une forte complémentarité peut être observée entre les rôles de l'information numérique et de l'information symbolique. Prenons l'exemple d'une carte et d'une image aérienne de la même région. L'information numérique portée par l'image permet d'obtenir une description assez précise de la scène, mais l'interprétation qui y est attachée est difficile à conclure. Une utilisation importante des représentations numériques dans la fusion de données consiste à quantifier l'incertitude ou la fiabilité de l'information (cette information peut être de nature numérique aussi bien que symbolique).

1.4.6 Imperfections de l'information

Les approches de fusion d'informations reposent sur une bonne modélisation des imperfections de l'information pour pouvoir en tenir compte le mieux possible [8, 43, 44]. En effet, à partir du moment où les informations sont imparfaites, trois actions s'offrent à nous :

— soit l'on cherche à les supprimer ;

— soit l'on les tolère en faisant en sorte que les algorithmes mis en jeu soient robustes face à ces imperfections ;

— soit l'on cherche à les modéliser.

C'est la troisième solution qui est retenue par les approches de fusion d'informations. L'information est relative soit à la source soit aux informations à combiner. L'information liée à la source peut être vue comme une information supplémentaire [8] venant s'ajouter aux informations à combiner. Elle est alors intégrée dans la modélisation. Les imperfections des informations à fusionner sont nombreuses. Les principales imperfections considérées sont l'incertitude et l'imprécision [8, 43, 45–47]. A ces deux imperfections peuvent être ajoutée l'incomplétude [8, 45, 46] et le conflit [8, 43, 47, 48].

Incertitude

L'incertitude caractérise le degré de conformité à la réalité d'une information, ou encore l'assurance d'une source selon l'information fournie. Une information incertaine décrit donc une connaissance partielle de la réalité, alors que ce qui est certain en-

traîne nécessairement la connaissance de toute la réalité. Pour mesurer l'incertitude, la réalité doit donc être connue. Par exemple : Demain, il fera chaud. Le fait que la réalité (le futur) ne soit pas connue, ne nous permet pas d'affirmer qu'il fera chaud demain. Cette information est donc incertaine. L'incertitude de l'information est l'imperfection la plus étudiée [8, 37]. Une des principales raisons de ces études provient du fait que les approches probabilistes permettent de modéliser parfaitement l'incertitude. L'incertitude correspondant à une ignorance (partielle ou totale) de la connaissance, peut être décrite à ce titre comme objective ou subjective [30]. L'objectivité de l'incertitude est liée à la description du monde et à l'information tandis que la subjectivité de l'incertitude est liée à l'opinion d'une source ou d'expert sur la vérité.

Imprécision

L'imprécision d'une information est caractérisée par le contenu de l'information. Elle est relative à l'information ou à la source. Elle mesure un défaut quantitatif de connaissance. Nous distinguons l'imprécision sans erreur (par exemple : il pleut beaucoup) de l'imprécision avec erreur (par exemple : il pleut 15 mm^3 par heure alors qu'il y en a 20) [30]. Dans le premier cas, l'imprécision est difficile à mesurer et reste relative à l'idée que la source s'en fait, alors que dans le second cas, elle se mesure par l'erreur de l'information fournie (défaut quantitatif de l'information). La réalité doit donc être connue ou estimée. L'imprécision est rencontrée dans de nombreux problèmes, ce qui a entraîné de nombreux développement de modèles de l'imprécision [8, 30].

Lien entre incertitude et imprécision

L'imprécision d'une information est souvent confondue avec l'incertitude [8, 34]. L'imprécision est donc un défaut quantitatif de la connaissance, alors que l'incertitude est un défaut qualitatif de la connaissance. De plus, la principale cause de l'incertitude d'une information provient de l'imprécision de l'information [8]. En effet, dans le cas d'informations quantitatives, l'imprécision d'une donnée entraîne une incertitude sur l'information véhiculée. Ainsi, l'incertitude peut induire l'imprécision.

Incomplétude

L'incomplétude est le manque d'information apporté par une source. L'incomplétude de l'information peut être la cause de l'incertitude et de l'imprécision. Elle peut se mesurer par la différence entre la quantité d'information réellement fournie par la source et la quantité d'information que la source doit fournir (en cas de bon fonctionnement ou pour répondre à un problème posé). La quantité d'information est cependant difficile à estimer. L'une des premières mesures de la quantité d'information a été introduite par C.E. Shannon [49], connue sous le nom d'entropie de Shannon.

Conflit

Deux informations ou plus sont en conflit si elles sont contradictoires. Le conflit peut aussi être rapporté à la source par extension. Par exemple une source sonar détecte une mine à un endroit, alors qu'une autre décrit une matière rocheuse. Les origines du conflit proviennent essentiellement de trois situations [8] :

— Les sources ne sont pas fiables;

— Le cadre de discernement est non exhaustif. L'hypothèse d'un monde fermé est alors fausse;

— Les sources observent des phénomènes totalement différents. Dans ce cas il ne faut pas combiner les informations.

1.4.7 Méthodes de fusion d'informations

Généralement, on rencontre deux (02) catégories de fusion : celles basées sur des modèles physiques et celles basées sur l'utilisation des techniques de l'intelligence artificielle et de la classification automatique qui peuvent se baser sur plusieurs théories comme la théorie des probabilités, la théorie de l'évidence, la théorie des possibilités et les sous-ensembles flous. Nous reprenons, le tableau comparatif (Tableau 1.2) des différentes techniques de fusion effectué par [34].

TABLE 1.2 – Analyse comparative des techniques de fusion [34]

Techniques de fusion	Avantages	Limites
Filtre de Kalman [50]	Mise en œuvre facile. Efficacité en termes de calcul.	Problèmes de précision. Trop de composants vectoriels aboutissant à un surcoût computationnel.
Filtre de Kalman étendu [51]	Plus efficace que le filtre Kalman. Fonctionne bien dans des problèmes d'incertitude.	Haute complexité de calcul de précision. Résultats instables en petites durées.
Méthodes de covariance [52]	Facile à utiliser et à mettre en œuvre. Précision des résultats.	Des comportements contradictoires dans certains cas.
Machines à vecteur de support (SVM [7]	Efficace pour les données de capteurs hétérogènes. De meilleurs résultats pour des données de capteurs incompatibles.	Sensible au bruit. Des problèmes de catégorisation dans certains cas.
Réseaux bayésiens [53]	Fonctionne bien lorsque des informations multi-sources sont disponibles. Peut combiner plusieurs sources de données.	Haute exigence de mémoire. Consomme plus de temps avant de produire un résultat.
La logique floue [54]	Méthodes basées sur la décision. Réponse rapide aux changements rapides.	Plus d'efforts de calcul.
Réseaux de neurones artificiels [6]	Traite des données non linéaires. Efficace pour les données de capteurs hétérogènes.	Haute complexité.
La théorie de Dempster-Shafer [55]	Méthode fondée sur la décision. Traite des problèmes statistiques et modélise les incertitudes.	La complexité augmente lorsque les capteurs augmentent. La croyance et la probabilité peuvent différer parfois.

1.5 Conclusion

Dans ce chapitre, nous avons présenté les notions clés afin de mener à bien nos travaux de recherche. Il s'agit des concepts de l'Internet des Objets, de la qualité de l'information et la fusion d'informations. D'abord, au niveau de l'Internet des Objets, nous avons présenté le concept proprement dit et les défis majeurs dont il fait cas. Ensuite, nous avons présenté la qualité de l'information. Des notions importantes des problèmes liés à la qualité de l'information ont été abordées. Nous avons mentionné également les facteurs liés à la qualité de l'information dans les systèmes distribués. Enfin, nous avons présenté le concept de la fusion d'information à travers des définitions, les différents imperfections ainsi que quelques théories utilisée pour la résolution des problèmes de fusion.

2

La Théorie de Dempster-Shafer

2.1 Introduction 34
2.2 Concepts de base 35
2.3 Défis de la théorie de Dempster-Shafer 42
2.4 Conclusion 43

2.1 Introduction

Dans la section 6.5 du chapitre 1, nous avons vu quelques théories de fusion d'information. Parmi ces méthodes, plusieurs chercheurs ont montré que la théorie de Dempster-Shafer a des atouts prometteurs pour résoudre les problèmes liés à la qualité de l'information en environnement IoT [2,33]. Cette théorie, inspirée de la théorie des probabilités supérieures et inférieures, repose sur la modélisation de la croyance en un évènement. Cette modélisation se réalise à partir de fonctions de masse permettant

une bonne représentation des connaissances. Cette théorie permet particulièrement de bien modéliser l'incertitude. Ce chapitre vise à présenter les différents aspects de cette théorie.

2.2 Concepts de base

La théorie de Dempster-Shafer, également connue comme la théorie de l'évidence, ou théorie des fonctions de croyance, est une théorie mathématique du raisonnement sur l'évidence et la plausibilité. Elle a été développée par Glenn Shafer [56] en se basant sur des travaux antérieurs de Arthur Dempster [57]. Elle a attiré l'attention des chercheurs en intelligence artificielle au début des années 80.

Dans un espace discret fini, la théorie de Dempster-Shafer représente la généralisation de la théorie des probabilités discrètes dans laquelle les probabilités sont assignées à des ensembles, contrairement aux singletons mutuellement exclusifs.

2.2.1 Quelques notations

Nous allons présenter ici les conventions de notation utilisées dans la suite de ce manuscrit. Soit Ω un univers constitué d'un ensemble fini qui contient toutes les propositions étudiées, appelé également cadre de discernement. L'ensemble qui ne contient aucun élément de Ω est représenté par $\varnothing$. $\mathcal{P}(\Omega)$ représente l'ensemble contenant toutes les parties (les sous-ensembles) de Ω. Les opérations binaires sur les ensembles sont l'inclusion $\subset$, l'union $\cup$ et l'intersection $\cap$. Une hypothèse est un élément de $\mathcal{P}(\Omega)$.

La théorie de Dempster-Shafer est caractérisée par trois fonctions principales : **l'élément de preuve** [1], **la croyance** [2] et **la plausibilité** [3].

2.2.2 Élément de preuve

Appelé aussi la fonction de masse, noté m ou **assignement de probabilité de base** [4], c'est une fonction qui affecte le degré d'évidence disponible à un élément A, et seule-

1. Body Of Evidence (*BOE*)
2. Belief
3. Plausibility
4. Basic probability assignement (*BPA*)

ment à A, de $\mathcal{P}(\Omega)$ dans l'intervalle $[0,1]$, tel que la somme des $m(A)$ est égale à 1.

Formellement on a:

$$m : \mathcal{P}(\Omega) \mapsto [0,1] \tag{2.1}$$

$$m(\varnothing) = 0 \tag{2.2}$$

$$\sum_{A \in \mathcal{P}(\Omega)} m(A) = 1 \tag{2.3}$$

$m(\varnothing)$ – représentant l'absence de solution est toujours égale à 0 car Ω doit être exhaustive. Chaque élément A tel que $m(A) > 0$ est appelé élément focal [58].

Généralement, le *BOE* n'est pas un équivalent de la fonction de probabilités classique. En effet, la valeur exacte de la probabilité (dans le sens classique) appartient à un intervalle borné par deux valeurs qui sont *la croyance* et *la plausibilité* [59].

Exemple 2.1

Dans une enquête sur un incident de vol, trois personnes *P1*, *P2*, *P3* sont accusées. L'enquêteur pense à 30% que $P1$ est le voleur, à 50% que l'un des deux autres personnes $P2$ ou $P3$, est le voleur, et à 10% que $P1$ ou $P3$ le soit.
Le cadre de discernement s'écrit ainsi :

$$\Omega = \{P1, P2, P3\}.$$

L'ensemble de parties de Ω est :

$$\mathcal{P}(\Omega) = \{\varnothing, \{P1\}, \{P2\}, \{P3\}, \{P1, P2\}, \{P1, P3\}, \{P2, P3\}, \Omega\}$$

La distribution des masses s'exprime ainsi :

$$m(\{P1\}) = 0.3, m(\{P2, P3\}) = 0.5, m(\{P1, P3\}) = 0.1$$

et

$$m(\Omega) = 1 - (0.3 + 0.5 + 0.1) = 0.1$$

à partir de l'équation 2.3.
Les masses des hypothèses restantes sont toutes égales à 0.

2.2.3 Croyance

La croyance est la quantité de confiance qui supporte une hypothèse A de $\mathcal{P}(\Omega)$, toutes parties comprises. Elle est notée par $Cr(A)$ ou plus souvent $Bel(A)$. Il est clair que $Bel(\varnothing)$ est nulle, et $Bel(\Omega)$ est toujours 1. Il faut noter que la croyance est une mesure non additive, c'est-à-dire que $Bel(A) + Bel(\Omega - A) \neq 1$.

Formellement on a :

$$Bel(\varnothing) = 0 \tag{2.4}$$

$$Bel(\Omega) = 1 \tag{2.5}$$

$$Bel(A) = \sum_{B/B\subseteq A} m(B) \tag{2.6}$$

$$\forall p, q \in \mathcal{P}(\Omega) \quad Bel(p \cup q) \geq Bel(p) + Bel(q) - Bel(p \cap q). \tag{2.7}$$

le BPA est obtenu avec la fonction inverse suivante:

$$m(A) = \sum_{B/B\subseteq A} (-1)^{|A-B|} Bel(B), \tag{2.8}$$

où $|A - B|$ est la différence en termes de cardinalité des deux ensembles.

Exemple 2.2

En continuité de l'exemple précédent, la croyance de chaque hypothèse est calculée comme suit :

TABLE 2.1 – Les degrés de croyance calculés à partir de la distribution de masses

Hypothèse A	Croyance $Bel(A)$
$\varnothing$	0
$\{P1\}$	0.3
$\{P2\}$	0
$\{P3\}$	0
$\{P1, P2\}$	$0.3 + 0 + 0 = 0.3$
$\{P1, P3\}$	$0.3 + 0 + 0.1 = 0.4$
$\{P2, P3\}$	$0 + 0 + 0.1 = 0.1$
Ω	1

2.2.4 Plausibilité

La plausibilité représente la croyance qu'il est possible d'affecter à une hypothèse A de $\mathcal{P}(\Omega)$; autrement dit, c'est la croyance affectée à A et aux hypothèses dont l'intersection avec A n'est pas vide. Elle est notée $Pl(A)$. La valeur de la plausibilité est toujours supérieure ou égale à celle de la croyance. Identiquement à la croyance, $Pl(\varnothing)$ est nulle, et $Pl(\Omega)$ vaut 1.

Formellement on a :

$$Pl(\varnothing) = 0 \tag{2.9}$$

$$Pl(\Omega) = 1 \tag{2.10}$$

$$Pl(A) = \sum_{B/B\cap A\neq\varnothing} m(B) \tag{2.11}$$

La plausibilité peut aussi être dérivée à partir de la croyance :

$$Pl(A) = 1 - Bel(\bar{A}), \tag{2.12}$$

où $\overline{A}$ est le complément de A dans Ω.

La probabilité classique d'un événement A, $P(A)$ est représentée par l'intervalle qui a comme bornes inférieure et supérieure $Bel(A)$ et $Pl(A)$, respectivement, c'est-à-dire

$$Bel(A) \leq P(A) \leq Pl(A)$$

Si

$$Bel(A) = Pl(A),$$

alors la probabilité de A est unique et

$$P(A) = Bel(A) = Pl(A)$$

Exemple 2.3

Les degrés de plausibilité de l'exemple précédent sont donnés par :

TABLE 2.2 – Les degrés de plausibilité calculées à partir de la distribution de masses

Hypothèse A	Plausibilité $Pl(A)$
$\varnothing$	0
$\{P1\}$	$0.3 + 0.1 + 0.1 = 0.5$
$\{P2\}$	$0.5 + 0.1 = 0.6$
$\{P3\}$	$0.5 + 0.1 + 0.1 = 0.6$
$\{P1, P2\}$	$0.3 + 0.5 + 0.1 + 0.1 = 1$
$\{P1, P3\}$	$0.3 + 0.5 + 0.1 + 0.1 = 1$
$\{P2, P3\}$	$0.5 + 0.1 + 0.1 = 0.7$
Ω	1

2.2.5 Règles de combinaison

Souvent, l'information est obtenue à partir de sources distinctes et indépendantes, et généralement ces sources ne sont pas parfaites. Il est donc nécessaire de fusionner les évidences fournies par ces sources en utilisant des règles de combinaison. Dans le modèle de Dempster-Shafer, il existe trois types de règles de combinaison : *conjonctive*, *disjonctive*, et *mixte*.

Combinaison conjonctive

Dempster a introduit une règle connue sous le nom de règle de combinaison de Dempster [5] pour fusionner deux pièces d'évidence qui peuvent être représentées par deux fonctions de croyance. Cette règle respecte les quatre propriétés algébriques suivantes : l'associativité, la commutativité, l'idempotence et la continuité [59].

Soient deux fonctions de masse $m_1(A)$ et $m_2(A)$ associées à deux fonctions de croyance $Bel_1(A)$ et $Bel_2(A)$ respectivement. $(Bel_1 \oplus Bel_2)(A)$ est définie à travers la masse $m_{1\oplus 2}(A)$:

$$m_{1\oplus 2}(A) = \sum_{B\cap C=A} m_1(B)m_2(C). \tag{2.13}$$

Néanmoins, cette règle a été critiquée par de nombreux chercheurs, parce qu'elle n'est pas normalisée. En d'autre termes, elle permet d'affecter des masses à des ensembles vides. Ce problème est dû au fait que l'intersection de deux éléments focaux peut générer $\varnothing$, ce qui dénote un conflit entre les croyances.

Pour corriger ce problème, Shafer a proposé d'ignorer carrément les conflits et

5. Dempster's Combination Rule

d'étendre la distribution des masses. Le facteur de conflit est une valeur $K < 1$ et définie par l'équation 2.14 :

$$K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C). \tag{2.14}$$

Afin de normaliser la distribution de masse, la règle de Dempster doit être multipliée par la quantité $(1 - K)^{-1}$. Donc la règle de Dempster-Shafer [56] est représentée par les équations 2.15 et 2.16 :

$$m(\varnothing) = 0 \tag{2.15}$$

$$m(A) = \frac{1}{1 - K} \sum_{B \cap C = A \neq \varnothing} m_1(B) m_2(C) \tag{2.16}$$

Cependant, la règle de Shafer n'est pas toujours satisfaisante à cause de la normalisation qui considère que le conflit vient de l'imperfection des sources.

D'autres chercheurs ont proposé des règles différents, comme Smets [60] qui a introduit une forme non normalisée pour prendre en compte les univers ouverts. Il suppose que le conflit vient du fait que l'ensemble Ω n'est pas exhaustif. Pour cette raison, la règle de Smets affecte une masse K à $\varnothing$ qui représente une hypothèse manquante. Elle est définie par les équations 2.17 et 2.18.

$$K = m(\varnothing) = \sum_{B \cap C = \varnothing} m_1(B) m_2(C) \tag{2.17}$$

$$m(A) = \sum_{B \cap C = A \neq \varnothing} m_1(B) m_2(C) \tag{2.18}$$

Par contre, Yager [61] propose un modèle d'un univers fermé (Ω est exhaustif). La mesure de conflit est ajoutée à la mesure du cadre de discernement Ω. Donc le conflit sera transformé en ignorance [8]. Ainsi :

$$m(\varnothing) = \sum_{B \cap C = \varnothing} m_1(B) m_2(C) \tag{2.19}$$

$$m(\Omega) = \left(1 - \sum_{A \in \mathcal{P}(\Omega)} \sum_{B \cap C = A \neq \varnothing} m_1(B) m_2(C)\right) + \sum_{B \cap C = \varnothing} m_1(B) m_2(C) \tag{2.20}$$

Combinaison disjonctive

Dans la combinaison disjonctive, l'union est utilisée à la place de l'intersection. Les éléments focaux sont obtenus à partir de la table d'union. La fonction de masse de la règle de combinaison disjonctive est formalisée par l'équation 2.21 :

$$m(A) = \sum_{B \cup C = A \neq \varnothing} m_1(B) m_2(C). \tag{2.21}$$

L'union de deux éléments focaux ne peut être jamais vide, donc que $m(\varnothing) = 0$. Cela veut dire qu'il est impossible d'obtenir un conflit lors de la combinaison. En contrepartie, l'union peut causer une perte de spécificité puisque les masses sont élargies. Cette approche apparaît plus intéressante en l'absence des informations nécessaires sur la fiabilité des sources.

Combinaison mixte

Dubois et Prade [62] ont proposé une combinaison mixte comme un compromis entre la combinaison conjonctive et la combinaison disjonctive comme un essai de conserver les avantages des deux [8]. De même que la combinaison de Yager, le modèle de Dubois et Prade suppose que le conflit est dû à la non fiabilité des sources. La règle de Dubois et Prade est définie par l'équation 2.22 :

$$m(A) = \sum_{B \cap C = A \neq \varnothing} m_1(B) m_2(C) + \sum_{\substack{B \cup C = A \neq \varnothing \\ B \cap C = \varnothing}} m_1(B) m_2(C) \tag{2.22}$$

2.2.6 Décision

Prendre une décision veut dire choisir une hypothèse élémentaire parmi les autres par la maximisation d'un critère. La sélection est établie en observant la croyance et la plausibilité de chaque hypothèse résultante. Il existe trois règles de décision : le *maximum de croyance*, le *maximum de plausibilité* et le *maximum de probabilité pignistique*

Maximum de croyance

Dans cette règle, il s'agit de l'hypothèse élémentaire ω_i qui a la croyance maximale. C'est-à-dire que ω_i est choisie si elle satisfait l'équation 2.23 :

$$Bel(\omega_i) = \max_{1 \leq j \leq n} Bel(\omega_j) \tag{2.23}$$

Cette méthode est connue par son caractère très pessimiste [8].

Maximum de plausibilité

En opposition à la règle précédente, l'élément d_i est sélectionné par le maximum de plausibilité. En d'autres termes, d_i est choisi de telle manière à résoudre l'équation 2.24 :

$$Pl(\omega_i) = \max_{1 \leq j \leq n} Pl(\omega_j) \tag{2.24}$$

Ainsi, cette méthode a un caractère très optimiste [63].

Maximum de probabilité pignistique

Cette règle est proposée par Smets et Kennes [64]. Elle consiste à transformer la fonction de masse $m(A)$ en une fonction de probabilité $BetP(\omega)$. Cette transformation est appelée *transformation pignistique* et elle est définie par l'équation 2.25.

$$BetP(\omega) = \sum_{\substack{A \subseteq \Omega \\ \omega \in A}} \frac{1}{|A|} \frac{m(A)}{1 - m(\varnothing)} \tag{2.25}$$

Dans cette transformation, la masse de croyance $m(A)$ est distribuée uniformément sur l'ensemble des éléments de A. Le critère est de prendre l'élément qui a le maximum de probabilité pignistique. Cette méthode est vue comme un compromis entre les deux précédentes.

2.3 Défis de la théorie de Dempster-Shafer

La théorie de Dempster-Shafer est beaucoup utilisée dans les problèmes de fusion d'informations [30, 43, 52]. Dans l'application de cette théorie, de nombreux défis sont

de plus en plus abordés. Nous pouvons citer entre autres, le problème de la quantification de l'incertitude [48,65,66] et le problème de la gestion des conflits entre différentes sources d'information [65, 67].

2.3.1 La quantification de l'incertitude

Dans la théorie des croyances, la question de savoir comment mesurer le degré d'incertitude des rapports des capteurs ou des sources est un problème ouvert [37, 48, 68]. L'une des principales problématiques est liées à prise en compte du cadre de discernement dans la quantification de l'incertitude. Plusieurs théories et méthodes ont été proposées. Ces théories se basent particulièrement sur l'entropie de Shannon [49] pour la quantification de l'incertitude. Dans la revue de littérature nous verrons plus en détail, quelques modèles proposés et leur limites.

2.3.2 Gestion des conflits

La théorie des croyances est une théorie robuste pour la fusion des informations multisources. Cependant l'un des problèmes cruciaux est celui de la gestion des conflits lorsqu'il existe des informations contradictoires multisources [33, 69]. Une meilleure estimation du conflit est un enjeu important permettant d'aboutir à la vérification de véracité des informations [70, 71].

2.4 Conclusion

Dans ce chapitre, nous avons mis en exergue le principe de base de la théorie de Dempster-Shafer. Cette théorie est aujourd'hui abordée dans de nombreuses études liées à la fusion d'information dans les systèmes distribués. Cependant, certains problèmes dans l'utilisation de cette théorie à savoir, la quantification de l'incertitude et la gestion des conflits demeurent ouverts.

3

Incertitudes et conflits de sources d'information dans les systèmes distribués

3.1 Introduction 44
3.2 Mesures d'incertitude 45
3.3 Méthodes de résolution de conflits 50
3.4 Conclusion 52

3.1 Introduction

Ce chapitre résume quelques méthodes utilisées pour améliorer la qualité de l'information dans l'Internet des Objets. Nous avons vu que dans l'Internet des Objets, les

applications s'appuient sur des éléments d'information recueillis auprès de capteurs de niveaux de fiabilité différents. Ces capteurs recueillent, génèrent et préservent une variété d'informations avec diverses représentations. Dans ce chapitre, nous montrons quelques travaux ayant abordé le problème de la qualité de l'information dans les systèmes distribués en général. Nous avons focalisé notre étude sur deux aspects majeurs notamment la quantification de l'incertitude ainsi que la véracité des informations. Ce chapitre est organisé comme suit. La section 3.2 présente quelques mesures basées sur la théorie de Dempster-Shafer. La section 3.3 présente des méthodes de résolution de conflit aussi connues sous l'appellation de méthode de découverte de véracité. La section 3.4 conclut ce chapitre.

3.2 Mesures d'incertitude

3.2.1 Quelques mesures d'incertitude dans la théorie de l'évidence de Dempster-Shafer

Dans le framework de Dempster-Shafer, l'on note plusieurs mesures d'incertitude. La plupart de ces mesures, le plus souvent sous une forme hybride, étendent le principe de l'entropie de Shannon pour mesurer l'incertitude.

a. Entropie de Shannon

En théorie de l'information, l'entropie de Shannon (E_s) [49] est utilisée pour mesurer le volume d'information dans un système, d'un processus ou d'un message. Cette mesure quantifie la valeur attendue de l'information contenue dans un message. La mesure est la suivante :

$$E_s(m) = -\sum_{i=0}^{n} p_i \log_b p_i \tag{3.1}$$

où n représente la quantité ou le nombre d'états de base, p_i représente la probabilité de l'état i satisfaisant $\sum_{i=0}^{n} p_i = 1$, b est la base du logarithme, elle prend le plus souvent, la valeur 2.

b. Mesure de confusion de Holhe

L'expression de la mesure de confusion de Hohle [72] est la suivante :

$$C_h(m) = -\sum_{A \subseteq X} m(A) \log_2 Bel(A) \tag{3.2}$$

c. Mesure de dissonance de Yager

L'expression de la mesure de dissonance de Yager [73] est la suivante :

$$E_y(m) = -\sum_{A \subseteq X} m(A) \log_2 Pl(A) \tag{3.3}$$

d. L'entropie pondérée de Hartley par Dubois & Prade

L'expression de la mesure de Hartley par Dubois & Prade [58] est la suivante :

$$E_{dp}(m) = -\sum_{A \subseteq X} m(A) \log_2 \mid A \mid \tag{3.4}$$

e. La mesure de discorde de Klir & Ramer

L'expression de la mesure de conflit de Klir & Ramer [74] est la suivante :

$$D_{kr}(m) = -\sum_{A \subseteq X} m(A) \log_2 \sum_{B \subseteq X} m(B) \frac{\mid A \cap B \mid}{\mid B \mid} \tag{3.5}$$

f. La mesure de conflit de Klir & Parviz

L'expression de la mesure de confusion de Klir & Parviz [75] est la suivante :

$$S_{kp}(m) = -\sum_{A \subseteq X} m(A) \log_2 \sum_{B \subseteq X} m(B) \frac{\mid A \cap B \mid}{\mid A \mid} \tag{3.6}$$

g. La mesure de conflit de George & Pal

L'expression de mesure de conflit de George & Pal [76] est la suivante :

$$C_{gp}(m) = -\sum_{A \subseteq X} m(A) \log_2 \sum_{B \subseteq X} m(B)[1 - \frac{\mid A \cap B \mid}{\mid A \cup B \mid}] \tag{3.7}$$

h. La mesure de Deng

L'expression de l'entropie de Deng [68] est la suivante :

$$E_d(m) = -\sum_{A \subseteq X} m(A) \log_2 (\frac{m(A)}{2^{|A|} - 1}) \tag{3.8}$$

i. La mesure de Zhou *et al.*

L'expression de la mesure de Zhou *et al.* [48] est la suivante :

$$E_z(m) = -\sum_{A \subseteq X} m(A) \log_2\left(\frac{m(A)}{2^{|A|}-1} e^{\frac{|A|-1}{|X|}}\right) \tag{3.9}$$

3.2.2 Autres modèles

a. Modèle de Smets

Smets [60] a proposé une taxonomie traitant le problème de l'incertitude. Il fait la distinction entre imprécision et incertitude. Dans sa vision, l'imprécision est une caractéristique du contenu de l'information, tandis que l'incertitude caractérise un manque d'information pour pouvoir décider si une affirmation est vraie ou fausse. Dans son étude, il donne l'exemple de deux affirmations :

— Noëlle a au moins deux enfants et j'en suis sûr.

— Noëlle a trois enfants, mais je n'en suis pas sûr.

Dans le premier cas, il s'agit d'une information imprécise mais certaine, tandis que dans le deuxième cas, l'information est précise mais incertaine. De cet exemple, on peut déduire que les deux concepts peuvent coexister, mais qu'ils ont des sens différents. Ainsi, il est possible d'avoir une information moins précise mais avec un niveau élevé de certitude ou d'avoir une information plus précise mais d'un niveau plus faible de certitude. Smets va plus loin dans son raisonnement et affirme qu'il est possible d'énoncer un principe de maximum d'information qui dit que « le produit » entre le niveau de précision et le niveau de certitude ne peut pas dépasser un certain niveau critique, comme dans le cas du principe d'incertitude de Heisenberg-Gabor utilisé dans le domaine du traitement du signal. L'incertitude peut être soit une propriété objective de l'information, soit une propriété subjective de l'observateur. L'incertitude objective peut être soit aléatoire, soit vraisemblable, si elle est connue statistiquement. L'incertitude subjective est dépendante de l'utilisateur : informations crédibles mais pas en totalité, informations de provenance douteuse ou informations non-pertinentes et dans ce dernier cas, elles ne sont pas utiles. Smets divise l'imprécision en deux catégories caractérisant : les informations sans erreurs et les informations avec erreurs. Les informations sans erreurs peuvent être vagues : pas bien définies et étant am-

bigües ; approximatives : bien définies et connues d'être proches de la vraie valeur ; floues : pas bien définies ; incomplètes : des valeurs manquantes ; déficientes : quand le manque de valeurs est important. Les informations avec erreurs peuvent être incorrectes : complètement erronées ; imprécises : la vraie valeur étant comprise dans un intervalle ; invalides : générant de fausses informations ; certaines informations peuvent être conflictuelles, incohérentes ou inconsistantes entres elles.

b. Modèle de Smithson

La taxonomie proposée par [77] s'appuie sur la notion d'ignorance ou celle de la non-connaissance. Smithson fait la distinction entre deux états différents : état d'ignorance et état d'ignorer. Quand une personne est ignorante à cause d'un manque de connaissance, elle est dans un état d'erreur par rapport à l'état de connaissance parfaite. Tandis que, dans le cas où une décision consciente est prise en déconsidérant quelque chose sous le motif de non-pertinence, cette information est ignorée. Ainsi, la notion d'ignorance est décomposée en deux grandes catégories : l'erreur et la non-pertinence. Smithson divise la non-pertinence en 3 catégories : l'inactualité, l'indécidabilité et le tabou. L'inactualité représente les choses qui n'ont plus de connexion avec la situation présente. L'indécidabilité est issue de problèmes qui n'ont pas de solution. Le tabou représente les situations dans lesquelles les utilisateurs n'ont pas le droit de poser de questions et ils doivent accepter ce qu'ils ont reçu. La grande majorité des études menées dans la modélisation de l'incertitude ont pris en compte l'erreur. Smithson identifie deux causes d'erreurs : l'incomplétude et la distorsion des données ou des informations. La distorsion des données/informations a pour cause principale l'imprécision, mais de possibles confusions peuvent également jouer un rôle important. L'incomplétude selon Smithson est due soit à l'absence de l'information, soit à l'incertitude de l'information. Afin de définir l'incertitude, Smithson propose de la subdiviser en trois catégories : la probabilité (l'incertitude des événements qui ne sont pas totalement prédictives) ; l'ambiguïté (l'incertitude provenant de l'incapacité à distinguer deux événements) et le flou (l'incertitude provenant des distinctions incorrectement observées).

c. Modèle de Bonissone et Tong

Bonissone et Tong [78] ont proposé, dans le cadre d'un système expert, un modèle

avec trois types d'ignorances : l'incertitude, l'incomplétude et l'imprécision. Dans leur vision l'incomplétude représente l'absence d'une valeur et l'imprécision caractérise une valeur qui n'est pas suffisamment précise. L'incertitude apparait lorsqu'un agent doit construire une opinion subjective sur une hypothèse, sans avoir une connaissance certaine. L'incomplétude, selon Bonissone et Tong, est définie lorsque l'inférence est réalisée en utilisant des informations partielles. Les éléments d'information sont ensuite regroupées en deux catégories : nécessaires et possibles. L'imprécision peut être représentée soit par un intervalle (ou par une valeur floue), soit par une disjonction ($Xest x_1 ou x_2$), soit encore par une négation (X *n'est pas* x_1). Malheureusement, Bonissone et Tong ne proposent pas de représentation mathématique particulière pour l'imprécision. Par contre, ils proposent de représenter l'incertitude soit par une probabilité (théorie des probabilités), soit par un degré de crédibilité (théorie de Dempster-Shafer), soit par une possibilité (théorie des possibilités).

d. Modèle de Lefebvre, Hadzagic et Bossé

Selon Lefebvre [79], les éléments principaux d'un système de fusion d'informations sont les sources d'information, le module de fusion et le résultat de la fusion. Pour chacun de ces trois éléments, des mesures de qualité peuvent être associées. La bonne détection des événements par les capteurs (les sources d'information) est influencée par des incertitudes. Pour caractériser la qualité des informations fournies par les sources d'information, [79] a utilisé la mesure de confiance. En fonction de cette mesure le module de fusion peut être construit. En ce qui concerne la qualité de l'information résultante du processus de fusion les mesures de complétude et de pertinence sont proposées.

e. Modèle de Rogova et Bossé

Rogova [29] a proposé une première démarche pour définir une ontologie de la qualité de l'information dans le cadre d'un processus de fusion d'information. Dans cette étude, trois catégories de qualité ont été définies : la qualité de la source d'information, la qualité du contenu de l'information et la qualité de la présentation de l'information. Ensuite, pour chaque catégories [29] a proposé plusieurs dimensions.

3.3 Méthodes de résolution de conflits

3.3.1 Méthodes basées sur la théorie des croyances

La théorie des preuves de Dempster-Shafer [56, 57] est aujourd'hui largement utilisé dans de nombreux domaines de fusion d'informations [68,80]. Toutefois, les résultats contre-intuitifs peuvent être obtenus lorsqu'il s'agit de preuves ou d'informations contradictoires [67, 81]. Si ce problème ne peut être résolu efficacement, cela limitera considérablement l'application de cette théorie. De nombreuses approches ont donc été proposées pour résoudre le problème [60, 80, 82].
En général, il existe principalement deux types de méthodologies. L'une consiste à modifier la règle de combinaison de Dempster et l'autre, à prétraiter les données. Concernant la méthode de combinaison, des travaux comme ceux de Yager [83] et Smets [64], résolvent le problème des résultats combinés contre-intuitifs par une redistribution de données contradictoires. Cependant, plusieurs auteurs ont montré que la modification de la règle de combinaison détruit généralement les bonnes propriétés telles que la commutativité et l'associativité. De plus, la modification de la règle de la combinaison de preuves est jugée déraisonnable si les résultats contre-intuitifs sont causés par une défaillance des capteurs/sources. Concernant la méthodologie de prétraitement, plusieurs chercheurs notamment Murphy [82] et Jiang *et al.* [80] ont prétraité les données et ont obtenu un effet idéal, basé sur la règle de combinaison de Dempster. Le modèle le plus récent est l'ECRM (Evidential Conflict Resolution Method) est celui de Walid *et al.* [33] présenté à la figure 3.1.

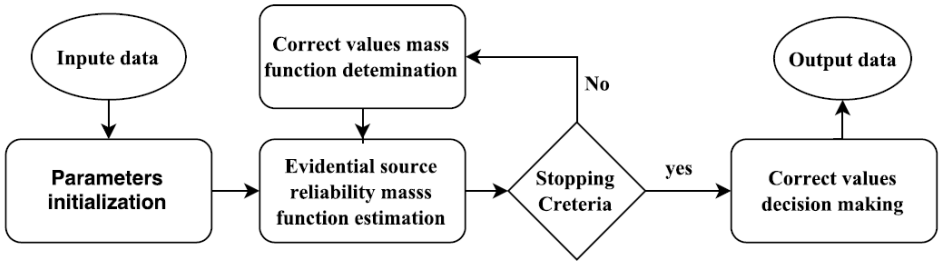

FIGURE 3.1 – Méthode ECRM [33]

La méthode ECRM se base sur la méthode de Murphy [82] pour résoudre le problème des informations contradictoires fournies par différentes sources dans les applications IoT. Il s'agit ici de déduire pour chaque source d'information les fonctions

de masse correctes qui permettront une bonne prise de décision. Le principe présenté à la figure 3.1 comprend trois principales (03) étapes : l'estimation de la fiabilité, la détermination des fonctions de masse correcte et la prise de décision sont basées sur la méthode de Murphy.

3.3.2 Méthodes de découverte de vérité

Le problème de la résolution des conflits [71] est aussi connu comme celui de la découverte de vérité [69, 84]. Ces méthodes sont classées en quatre catégories [69].

— Les méthodes basées sur les accords : Elles sont principalement liées au comptage du nombre de sources en accord ou en désaccord pour chaque information. Parmi ces méthodes, les plus connues sont TruthFinder [85] et 2-Estimates [38]. L'utilisation de ces algorithmes de découverte de vérité a permis de résoudre de nombreux problèmes liés aux conflits d'information multi-sources, surtout dans les domaines liés à la recherche d'information sur le Web. L'algorithme 1 présente la méthode TruthFinder.

Algorithme 1 : TruthFinder(S, D, V, ρ, γ, δ)

initialization. $\forall s \in S : T_s \longleftarrow 0.8$;
while *NonConvergence*(T_s, δ) **do**
  foreach $d \in D$ **do**
    foreach $v \in V_d$ **do**
      $\sigma_v \longleftarrow -\sum_{s \in S_v} \ln(1 - T_s)$;
      $\sigma_v^* \longleftarrow \sigma_v - \rho \sum_{v' \in V_d} \sigma'_v.sim(v, v')$;
      $C_v \longleftarrow \frac{1}{1+e^{-\gamma \sigma_v^*}}$;
  foreach $s \in S$ **do**
    $T_s \longleftarrow \frac{1}{|V_s|} \sum_{v \in V_s} C_v$;

Parmi les méthodes de découverte de vérité, TruthFinder proposée par Yin *et al.* [85] calcule la confiance d'une déclaration comme la probabilité de la véracité d'une source. L'idée principale qui sous-tend cette méthode est que la déclaration finale est susceptible d'être vraie si elle est fournie par des sources dignes de confiance tel que présenté dans l'algorithme 1. Cette méthode a été la première à incorporer la notion de similarité pour améliorer le décompte des voix

de ses valeurs similaires. Il utilise une fonction pondérée avec un paramètre de commande ρ et un facteur d'amortissement γ. Dans cette méthode, la véracité de chaque source, valeur comprise entre 0 et 1, est initialisée à 0.8. Truthfinder calcule pour chaque source, un score de confiance σ_v. La véracité est déduite par la moyenne $T_s = \sum_{v \in V_s} \frac{C_v}{|Vs|}$. ThruthFinder ajuste le score de confiance d'une réclamation afin qu'il intègre l'influence que des déclarations similaires peuvent avoir l'une sur l'autre à travers l'expression $\sigma_v^* \longleftarrow \sigma_v - \rho \sum_{v' \in V_d} \sigma_v'.sim(v, v')$. La confiance finale est calculée à l'aide de la fonction $C_v \longleftarrow \frac{1}{1+e^{-\gamma \sigma_v^*}}$.

Comme TruthFinder, 2-Estimates calcule la fiabilité de la source en faisant la moyenne des valeurs [69]. Cette méthode prend en considération les sources en désaccord pour chaque élément de données tout en calculant la confiance en fonction de l'accord et du désaccord des sources qui affirment des valeurs différentes. Cette méthode effectue une normalisation du décompte des voix et de la fiabilité de la source dans la plage de [0,1]. Comme condition d'arrêt, les auteurs ont utilisé un calcul de point fixe pour tester la convergence.

— Les méthodes basées sur l'estimation d'un maximum à postériori [19, 71] : ces méthodes sont basées sur d'éventuelles observations. Elles permettent de calculer les variables latentes optimales notamment la fiabilité ou la véracité de la source;

— Les méthodes analytiques [60, 84] : elles utilisent principalement, la diagonalisation matricielle pour la découverte de la vérité qui se pose comme un problème d'optimisation basé sur les graphes ;

— Les méthodes basées sur l'inférence bayésienne [71,78] : ces méthodes sont principalement basées sur la modélisation probabiliste bayésienne pour calculer la précision de la source.

3.4 Conclusion

Dans ce chapitre, nous avons présenté quelques approches permettant de résoudre les problèmes liés à la qualité de l'information dans les systèmes distribués. Deux groupes d'approches ont été étudiés. Il s'agit des approches pour la quantification de l'incertitude et des méthodes de résolution de conflits. Nous avons premièrement, présenté les mesures d'incertitude existantes et basées sur la théorie de Dempster-Shafer. Deuxièmement, nous avons étudié les méthodes de résolution de conflits. Pour cela,

en plus des méthodes basées sur la théorie de Dempster-Shafer, les méthodes de découverte de vérité ont été présentées.

4

Proposition d'une nouvelle mesure de l'incertitude

4.1 Introduction 54
4.2 Analyse et formulation du problème 56
4.3 Nouvelle mesure de l'incertitude 59
4.4 Exemples numériques et discussions 61
4.5 Conclusion 66

4.1 Introduction

L'enjeu majeur dans l'utilisation des applications *IoT*, consiste à recueillir, stocker et analyser les informations présentes dans l'environnement, à partir de dispositifs physiques (capteurs, experts, ...) et/ou virtuels (applications, sites web, ...),

pour lancer des actions ou aider à la prise de décision. La modélisation à travers les techniques de fusion de données est une question importante dans ces applications car plusieurs facteurs liés à l'incertitude (données bruitées, erronées, manquantes, ...) peuvent entraver la prise de décision. Ces dernières années, on assiste de plus en plus à des améliorations dans la gestion des problèmes d'incertitude dans les systèmes intelligents [4, 36, 37, 81]. Plusieurs théories notamment l'entropie de Shannon [49], la théorie des probabilités [31], la théorie des possibilités [62] et la théorie de Dempster-Shafer(TDS) [56,57] ont été développées. La théorie de Dempster-Shafer en particulier, a été largement utilisée dans diverses applications notamment la fusion d'information [86], l'aide à prise de décision [87], le diagnostic et la détection des fautes [35], la reconnaissance de cibles [88], etc.

Dans le modèle de Dempster-Shafer, on note plusieurs mesures d'incertitude hybrides ou étendues. Ces mesures utilisent le principe de l'entropie de Shannon pour la mesure de l'incertitude. Parmi ces mesures, on peut citer la mesure de confusion de Hohle [72], la dissonance de Yager [73], la mesure pondérée de Hartley [58], la mesure de Klir et Ramer [74], la mesure de Klir en Parviz [75], l'entropie de Deng [68] et l'entropie proposée par Zhou *et al.* [48].

Dans ce travail, nous nous intéressons principalement aux mesures de Deng et de Zhou *et al.*. La mesure de Deng, mesure très efficace, comparativement à diverses mesures dans certains cas à été utilisée dans plusieurs domaines d'application [68,89]. Cependant, l'une des principales limites de cette mesure d'incertitude est la non prise en compte de l'échelle du cadre de discernement (appelé FOD). Pour combler cette insuffisance, certains auteurs notamment Zhou *et al.*, ont proposé une mesure modifiée de Deng [48]. Néanmoins, bien qu'efficace dans certains cas, le problème lié au FOD est toujours perceptible car on assiste parfois à des mesures contre-intuitives. Dans ce manuscrit, nous proposons une nouvelle mesure d'incertitude en étendant les mesures de Deng et de Zhou *et al.*. Cette nouvelle mesure améliore la performance des mesures de Deng et de Zhou *et al.*, en prenant en considération l'ensemble des parties du FOD. Nous montrons après analyse des différentes méthodes existantes, la performance et l'efficacité de notre approche proposée.

Le présent chapitre est organisé comme suit: la section 4.2 présente le problème à résoudre. Dans la section 4.3, nous présentons la nouvelle mesure de l'incertitude. La section 4.4 présente, à l'aide d'exemples numériques, l'efficacité de la nouvelle mesure.

La conclusion ainsi que quelques perspectives liées à ce travail sont présentées à la section 4.5.

4.2 Analyse et formulation du problème

Comme mentionné dans l'état de l'art, plusieurs théories pour mesurer l'incertitude dans les environnements distribués ont été développées. Parmi ces théories, la théorie de Dempster-Shafer fournit des outils efficaces pour modéliser et traiter l'information incertaine [90]. Elle a été largement utilisée dans diverses applications notamment la fusion d'information [86], l'aide à la prise de décision [87], le diagnostic et la détection des fautes [35], la reconnaissance de cibles [88], etc.
Dans le modèle de Dempster-Shafer, on note plusieurs mesures d'incertitude. Parmi ces mesures, les plus récentes faisant l'objet de nombreuses applications sont l'entropie de Deng [68] et l'entropie proposée par Zhou *et al.* [48]. Dans ce travail, nous nous intéressons principalement à ces deux mesures d'incertitude.

4.2.1 Limites de l'entropie de Deng

Dans la théorie de Demspter-Shafer, l'information incertaine ne doit pas seulement être modélisée par les fonctions de masse, le *FOD* est aussi une source d'incertitude. Nous analysons ce problème en reprenant l'illustration de Zhou *et al.* [48]. Dans cet exemple, nous calculons les mesures de Deng (E_d) et de Zhou *et al.* (E_z).

Exemple 4.1

Soient les éléments de preuve (appelés BOEs) m_1 et m_2, représentant respectivement, les résultats de deux capteurs fiables dans un système de traitement d'informations. On a :

$$m_1 : m_1(\{a,b\}) = 0.4, m_1(\{c,d\}) = 0.6$$
$$m_2 : m_2(\{a,b\}) = 0.4, m_2(\{a,c\}) = 0.6$$

La mesure d'incertitude de Deng est la suivante :

$$\begin{aligned} E_d(m_1) &= -\sum_{A\subseteq X} m_1(A)\log_2\left(\frac{m_1(A)}{2^{|A|}-1}\right) \\ &= -0.4\times\log_2\left(\frac{0.4}{2^2-1}\right) - 0.6\times\log_2\left(\frac{0.6}{2^2-1}\right) \\ E_d(m_1) &= 2.5559 \end{aligned} \tag{4.1}$$

$$\begin{aligned} E_d(m_2) &= -\sum_{A\subseteq X} m_2(A)\log_2(\tfrac{m_2(A)}{2^{|A|}-1}) \\ &= -0.4\times\log_2(\tfrac{0.4}{2^2-1}) - 0.6\times\log_2(\tfrac{0.6}{2^2-1}) \\ E_d(m_2) &= 2.5559 \end{aligned} \tag{4.2}$$

Comme nous pouvons le constater, malgré la différence des *FODs* ($i.e. X_1 = \{a,b,c,d\}$ et $X_2 = \{a,b,c\}$) les résultats issus de la mesure de Deng concernant le *BOE* m_1 est identique à la mesure de l'incertitude de le *BOE* m_2. Intuitivement, la mesure d'incertitude du *BOE* m_1 devait être plus élevée que celle du *BOE* m_2. A cette principale limite, certains auteurs notamment Zhou *et al.* [48] ont proposé une mesure étendue de la mesure de Deng.
La mesure d'incertitude de Zhou *et al.* est la suivante :

$$\begin{aligned} E_z(m_1) &= -\sum_{A\subseteq X} m_1(A)\log_2(\tfrac{m_1(A)}{2^{|A|}-1}e^{\frac{|A|-1}{|X|}}) \\ &= -0.4\times\log_2(\tfrac{0.4}{2^2-1}e^{\frac{2-1}{4}}) - 0.6\times\log_2(\tfrac{0.6}{2^2-1}e^{\frac{2-1}{3}}) \\ E_z(m_1) &= 2.1952 \end{aligned} \tag{4.3}$$

$$\begin{aligned} E_z(m_2) &= -\sum_{A\subseteq X} m_2(A)\log_2(\tfrac{m_2(A)}{2^{|A|}-1}e^{\frac{|A|-1}{|X|}}) \\ &= -0.4\times\log_2(\tfrac{0.4}{2^2-1}e^{\frac{2-1}{3}}) - 0.6\times\log_2(\tfrac{0.6}{2^2-1}e^{\frac{2-1}{3}}) \\ E_z(m_2) &= 2.0750 \end{aligned} \tag{4.4}$$

Comme nous pouvons le remarquer, la mesure de Zhou *et al.* donne des résultats pratiquement différents pour chaque *BOE*. Dans cette mesure, nous pouvons constater que le degré d'incertitude calculé à partir des différents *BOEs* est réduit par rapport à la mesure de Deng. Cependant, bien que la mesure de Zhou *et al.* prenne en compte le nombre d'éléments dans le *FOD*, cette mesure donne souvent des mesures contre-intuitives.

4.2.2 Limite de l'entropie proposée par Zhou *et al.*

La principale limite de la mesure proposée par Zhou *et al.* se perçoit lorsque le *FOD* ne varie pas. Pour montrer cette limite, nous illustrons à partir de l'exemple 4.2, deux *BOEs* m_3 et m_4 représentant respectivement un cas d'incertitude totale et un cas de fonction de masse bayésienne.

Exemple 4.2

Considérons les BOEs m_3 et m_4 dans le cadre de discernement $X = \{a, b\}$. On a:

$$m_3 : m_3(\{a, b\}) = 1$$
$$m_4 : m_4(\{a\}) = m_4(\{b\}) = 0.5$$

La mesure de Zhou *et al.* est la suivante :

$$E_z(m_3) = -1 \times \log_2(\frac{1}{2^2-1} e^{\frac{2-1}{2}}) = 0.8636$$
$$E_z(m_4) = -0.5 \times \log_2(\frac{0.5}{2^1-1} e^{\frac{1-1}{2}}) - 0.5 \times \log_2(\frac{0.5}{2^1-1} e^{\frac{1-1}{2}}) = 1$$

La mesure de Deng est la suivante :

$$E_d(m_3) = -1 \times \log_2(\frac{1}{2^2-1}) = 1.5850$$
$$E_d(m_4) = -0.5 \times \log_2(\frac{0.5}{2^1-1}) - 0.5 \times \log_2(\frac{0.5}{2^1-1}) = 1$$

Comme nous pouvons le remarquer, le degré d'incertitude du *BOE* m_3 doit être supérieur à celui du *BOE* m_4, ce qui n'est pas le cas avec l'entropie de Zhou *et al.* (E_z). Par contre, la mesure de Deng (E_d) distingue mieux l'incertitude totale avec une mesure d'incertitude du *BOE* m_3 plus élevée que celle du *BOE* m_4.

4.2.3 Analyse des résultats

Après observation des exemples 4.1 et 4.2, on remarque que l'exemple 4.1 présente un cas de *FOD* variable avec le même nombre d'éléments dans l'ensemble des éléments focaux et l'exemple 4.2 présente un cas où le *FOD* ne varie pas. Dans l'exemple 4.1, la mesure proposée par Zhou *et al.* présente des résultats meilleurs comparativement aux résultats de Deng. Par contre, dans l'exemple 4.2 où on dispose d'un *FOD* fixe, la mesure proposée par Deng présente des résultats bien meilleurs comparativement aux résultats de Zhou *et al.*. De ce fait, comment quantifier de façon optimale l'incertitude en prenant en compte les limites observées dans les mesures proposées par Deng et Zhou *et al.*?

Pour apporter une solution à ce problème, nous proposons une nouvelle mesure d'incertitude en étendant les mesures de Deng et Zhou *et al.*.

4.3 Nouvelle mesure de l'incertitude

La particularité de la nouvelle mesure de l'incertitude est qu'elle prend en compte l'échelle de l'ensemble des parties du *FOD* notée 2^X. La nouvelle mesure de l'incertitude est définie comme suit :

$$E_{Nm}(m) = -\sum_{A \subseteq X} m(A) \log_2 \left(\frac{m(A)}{2^{|A|}-1} e^{\frac{|A|-1}{2^{|X|}}}\right) \tag{4.5}$$

où m est la fonction de masse définie sur le cadre de discernement X. A est l'élément focal de X et $\mid A \mid$ représente la cardinalité de A. La nouvelle mesure prend en compte le nombre d'éléments de l'ensemble des parties du *FOD* représenté par $2^{|X|}$, ce qui ne se perçoit pas dans les entropies de Deng et Zhou *et al.*.

Effectuons une simple transformation de la nouvelle entropie. On a :

$$E_{Nm}(m) = \sum_{A \subseteq X} m(A) \log_2 (2^{|A|}-1) - \sum_{A \subseteq X} m(A) \log_2 m(A) - \sum_{A \subseteq X} m(A) \log_2 e^{\frac{|A|-1}{2^{|X|}}} \tag{4.6}$$

Les deux premiers termes font référence à l'entropie de Deng [68]. Il s'agit respectivement de la mesure de la non-spécificité totale de la fonction de masse m, et de la mesure de la discorde de la fonction de masse entre divers éléments focaux. Le troisième terme, le facteur exponentiel, $e^{\frac{|A|-1}{2^{|X|}}}$, constitue le facteur principal de cette contribution. Le choix de ce facteur a été inspiré du facteur $e^{\frac{|A|-1}{|X|}}$ proposé par Zhou *et al.* [48], représentant la mesure de l'information incertaine dans un *BOE*. Comparativement à la mesure proposée par Zhou *et al.*, la nouvelle mesure prend en compte le nombre d'éléments dans l'ensemble des parties représenté par $2^{|X|}$. Le choix de l'ensemble des parties est lié à la définition du *BOE* dans le chapitre 2 selon laquelle une hypothèse appartient à l'ensemble des parties du *FOD*.

Ainsi, la nouvelle mesure se veut plus générique dans résolution des limites des mesures proposées par Deng et Zhou *et al.* évoquées respectivement dans les exemples 4.1

Dans la sous-section 4.4.1, la compatibilité de la nouvelle entropie de croyance avec l'entropie de Shannon (E_s), l'entropie de Deng (E_d) et l'entropie de Zhou *et al.* (E_z) est vérifiée à l'aide d'exemples énoncés par Deng et Zhou *et al.*. La sous-section 4.4.2 présente la supériorité de la nouvelle entropie des nouvelles croyances par rapport à d'autres mesures d'incertitude.

4.4.1 Concordance avec différentes mesures d'incertitude

La nouvelle mesure de l'incertitude que nous proposons est compatible aux entropies de base notamment l'entropie de Shannon (Equation 4.9) lorsque nous avons une fonction de masse bayésienne (c'est à dire un seul élément dans l'ensemble des éléments focaux).

$$\begin{aligned} E_{Nm}(m) &= -\sum_{A\subseteq X} m(A)\log_2\left(\frac{m(A)}{e^{|A|}-1}2^{\frac{|A|-1}{2^{|X|}}}\right) \\ &= -\sum_{A\subseteq X} m(A)\log_2\left(\frac{m(A)}{2^1-1}e^{\frac{1-1}{2^{|X|}}}\right) \\ &= -\sum_{A\subseteq X} m(A)\log_2 m(A) \\ E_{Nm}(m) &= E_s(m) \end{aligned} \tag{4.9}$$

Par ailleurs, une autre preuve de concordance de la nouvelle entropie avec différentes mesures d'incertitude est énoncée au cas où FOD dispose d'un seul élément comme l'indique l'exemple 4.3.

Exemple 4.3

Considérons un système de traitement d'information dans lequel l'information I rapportée par un capteur a une croyance égale 1. Dans le cadre du discernement $X = \{I\}$, la fonction de masse peut être ainsi notée $m(\{I\}) = 1$.

Le calcul des entropies de Shannon (E_s), Deng (E_d), Zhou (E_z) ainsi que la nouvelle entropie (E_{Nm}) est :

$$\begin{aligned} E_s(m) &= -1\times\log_2 1 = 0 \\ E_d(m) &= -1\times\log_2\frac{1}{2^1-1} = 0 \\ E_z(m) &= -1\times\log_2\left(\frac{1}{2^1-1}e^{\frac{1-1}{1}}\right) = 0 \\ E_{Nm}(m) &= -1\times\log_2\left(\frac{1}{2^1-1}e^{\frac{1-1}{2^1}}\right) = 0 \end{aligned}$$

Comme on peut le remarquer, dans un cas de certitude totale, la mesure d'incertitude

de la nouvelle mesure (E_{Nm}) est nulle de même que, les mesures de Shannon, Deng et Zhou.

Exemple 4.4

On considère un cadre de discernement $X = \{a, b, c\}$, *pour une fonction de masse* $m(a) = m(b) = m(c) = 1/3$.

Les entropies de Shannon (E_s), Deng (E_d), Zhou (E_z) ainsi que la nouvelle entropie (E_{Nm}) sont :

$E_s(m) = -1/3 \times \log_2 1/3 - 1/3 \times \log_2 1/3 - 1/3 \times \log_2 1/3 = 1.5850$

$E_d(m) = -1/3 \times \log_2 \frac{1/3}{2^1-1} - 1/3 \times \log_2 \frac{1/3}{2^1-1} - 1/3 \times \log_2 \frac{1/3}{2^1-1} = 1.5850$

$E_z(m) = -1/3 \times \log_2(\frac{1/3}{2^1-1} e^{\frac{1-1}{1}}) - 1/3 \times \log_2(\frac{1/3}{2^1-1} e^{\frac{1-1}{1}}) - 1/3 \times \log_2(\frac{1/3}{2^1-1} e^{\frac{1-1}{1}}) = 1.5850$

$E_{Nm}(m) = -1/3 \times \log_2(\frac{1/3}{2^1-1} e^{\frac{1-1}{2^1}}) - 1/3 \times \log_2(\frac{1/3}{2^1-1} e^{\frac{1-1}{2^1}}) - 1/3 \times \log_2(\frac{1/3}{2^1-1} e^{\frac{1-1}{2^1}}) = 1.5850$

De toute évidence, les exemples 4.3 et 4.4 ont montré que les résultats de l'entropie de Shannon, de l'entropie de Deng et de l'entropie de Zhou *et al.* sont identiques lorsque la croyance n'est attribuée qu'à un seul élément, ce qui est prouvé par la cohérence des probabilités de la nouvelle mesure proposé.

4.4.2 Supériorité de la nouvelle mesure de l'incertitude

Cette sous-section montre à l'aide d'exemples numériques, l'efficacité de la nouvelle mesure proposée par rapport aux autres mesures d'incertitude dans le cadre de Dempster-Shafer. Pour montrer la supériorité de la nouvelle mesure de l'incertitude, nous reprenons l'exemple de Deng [68]. Cet exemple décrit un cas de *FOD* X fixe de 15 éléments. Parmi les éléments focaux, on fait varier l'ensemble d'un élément focal. Dans cet exemple, il s'agit de l'élément focal X.

Exemple 4.5

Considérons une fonction de masse m *dans un FOD* X *telle que :* $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$, $m(\{6\}) = 0.05$, $m(\{3, 4, 5\}) = 0.05$, $m(E) = 0.8$ *et* $m(X) = 0.1$. E *représente un sous-ensemble d'éléments variants de 1 à 14 comme le montre le tableau 4.4.*

L'entropie de Deng, celle proposée par Zhou *et al.* ainsi que la nouvelle mesure de l'entropie sont calculées en fonction du nombre d'éléments dans le sous-ensemble E. Les résultats sont affichés dans le tableau 4.4 et représentés dans les figures 4.1, 4.2 et 4.3.

A ce niveau, on peut remarquer que dans cet exemple, montrant un cas de *FOD* qui ne varie pas comme dans l'exemple 4.2. La nouvelle mesure ne s'éloigne donc pas des mérites de la mesure de Deng.

TABLE 4.4 – Mesures d'entropie avec un sous-ensemble E d'éléments variables

Eléments dans E	Entropie de Deng [68]	Entropie de Zhou *et al.* [48]	Nouvelle mesure
$E = \{1\}$	2.6623	2.5180	2.6622
$E = \{1, 2\}$	3.9303	3.7090	3.9302
$E = \{1, ..., 3\}$	4.9082	4.6100	4.9080
$E = \{1, ..., 4\}$	5.7878	5.4127	5.7876
$E = \{1, ..., 5\}$	6.6256	6.1736	6.6254
$E = \{1, ..., 6\}$	7.4441	6.9151	7.4439
$E = \{1, ..., 7\}$	8.2532	7.6473	8.2530
$E = \{1, ..., 8\}$	9.0578	8.3749	9.0575
$E = \{1, ..., 9\}$	9.8600	9.1002	9.8597
$E = \{1, ..., 10\}$	10.6612	9.8244	10.6608
$E = \{1, ..., 11\}$	11.4617	10.5480	11.4613
$E = \{1, ..., 12\}$	12.2620	11.2714	12.2616
$E = \{1, ..., 13\}$	13.0622	11.9946	13.0617
$E = \{1, ..., 14\}$	13.8622	12.7177	13.8617

Comme nous pouvons le remarquer dans la figure 4.1, la nouvelle entropie et celle de Zhou *et al.* augmentent lorsque la taille du sous-ensemble E varie. Cependant les valeurs de la nouvelle entropie sont nettement supérieures à celle de Zhou *et al.*. Cela s'explique par le fait que la nouvelle entropie prend plus d'informations dans le *BOE* que celle de Zhou *et al.* qui enregistre des pertes d'informations surtout dans un tel cas

où l'on dispose d'un *FOD* qui ne varie pas, comme dans l'exemple 4.2.

Par ailleurs, la nouvelle entropie donne des résultats presque identiques à la mesure de Deng (figure 4.2) ce qui montre l'efficacité de l'approche lorsque le *FOD* ne varie pas. La nouvelle mesure ne s'éloigne pas des performances de la mesure de Deng.

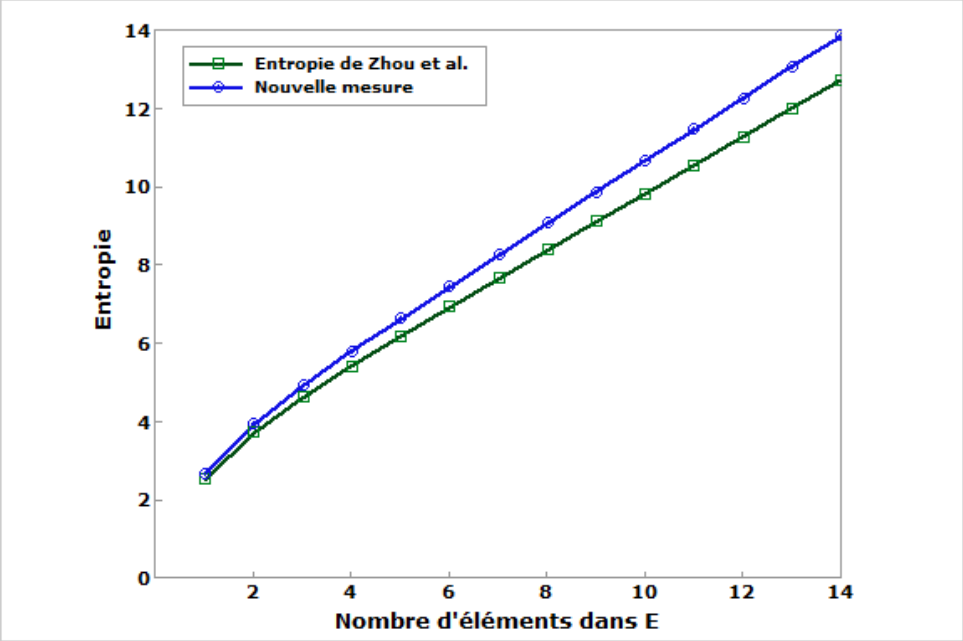

FIGURE 4.1 – Comparaison entre l'entropie de Zhou *et al.* et l'entropie modifiée

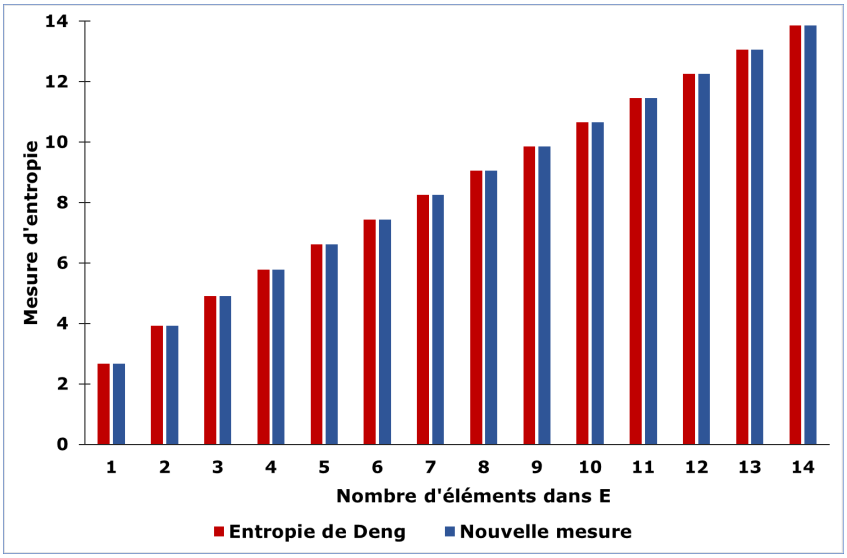

FIGURE 4.2 – Comparaison entre l'entropie de Deng et la nouvelle entropie

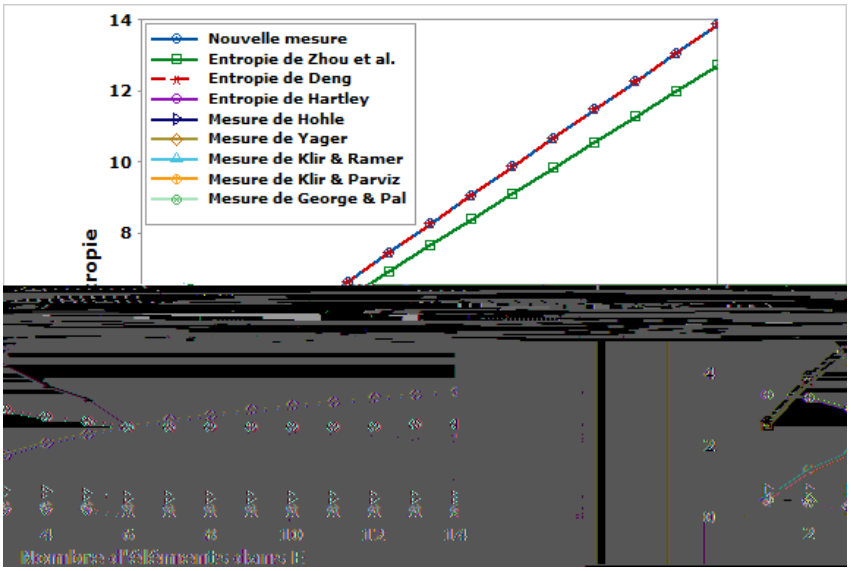

FIGURE 4.3 – Comparaison de la nouvelle mesure avec d'autres mesures d'incertitude

La figure 4.3 montre le degré d'incertitude du *BOE* en utilisant d'autres mesures d'incertitude différentes notamment la mesure de confusion de Hohle [72], la mesure de dissonance de Yager [73], la discorde de Klir & Ramer [74], la discorde de Klir & Parviz [75],la mesure de Dubois & Prade [58], la mesure de conflit de George & Pal [76], l'entropie de Deng [68] et l'entropie de Zhou *et al.* [48]. Dans cette figure, seules les entropies de Deng, Zhou *et al.*, Hartley et la nouvelle entropie augmentent de façon monotone avec l'augmentation de la taille du sous-ensemble E. Cependant avec l'augmentation de la taille de E, on observe soit une déclinaison, soit un changement de l'allure des autres mesures d'incertitude.

4.5 Conclusion

La quantification de l'incertitude dans les systèmes d'information est très importante pour l'évaluation de la qualité de l'information. Dans la littérature plusieurs méthodes basées sur l'entropie des croyances ont été proposées, mais celles - ci donnent des résultats contre-intuitifs, surtout en cas de *FOD* variable avec le *BOE*. Nous avons proposé une nouvelle mesure qui répond à ces insuffisances. Cette mesure, générique étend les mesures proposées par Deng et Zhou *et al.*. A partir d'exemples numériques et de propriétés mathématiques, nous avons montré l'efficacité de la nouvelle mesure.

5

Une approche basée sur l'entropie des croyances pour la résolution de conflits en environnement IoT

5.1 Introduction 68
5.2 Motivations et formulation du problème 69
5.3 Notre approche de résolution de conflit (ABC_IoT) 70
5.4 Expérimentations 75
5.5 Conclusion 85

5.1 Introduction

La plus grande force du paradigme de l'Internet des Objets est incontestablement l'impact qu'elle a sur la vie quotidienne des personnes. Ses applications reposent sur les informations prélevées par des sources d'information qui collectent, et génèrent une variété d'informations [91]. La fusion de données joue un rôle important dans ce nouveau concept. Elle consiste à combiner différentes informations issues de plusieurs sources en vue de faciliter la prise décision. Cette technique s'avère selon les etudes [3,37,92] beaucoup plus avantageuse qu'utiliser une seule source pour la prise de décision.

Cependant, dans un tel système où il y a des échanges d'informations, plusieurs problèmes liés à la qualité de l'information rendent complexe la fusion de données [2] pour assurer une aide à la décision efficace. Parmi ces problèmes, nous pouvons citer la gestion des informations contradictoires [71,84,93] qui soulève la question suivante : Comment obtenir une meilleure information si plusieurs sources sont en conflit?

Plusieurs travaux ont abordé ce problème [38,71,85,93]. Cependant, l'amélioration de la précision en vue d'aboutir à une meilleure prise de décision demeure toujours une issue ouverte dans les environnements intelligents. Pour apporter une ébauche de solution à ce problème, nous proposons une approche basée sur la théorie de Dempster-Shafer pour la résolution des conflits dans l'IoT (ABC_IoT). L'ABC_IoT utilise deux éléments fondamentaux notamment la quantification des incertitudes et la fiabilité des sources d'information.

Après cette introduction, la section 5.2 présente nos motivations et le formalisme du problème. Dans la section 5.3, nous présentons l'ABC_IoT, notre approche pour résoudre le problème de conflit en environnement IoT. La section 5.4 présente les expérimentations faites pour montrer l'efficacité de l'ABC_IoT. La section 5.5 conclut ce chapitre.

5.2 Motivations et formulation du problème

Une application IoT est caractérisée par son autonomie dans la prise de décision. Cette prise de décision se fait en fonction des informations collectées et fusionnées dans le système. La principale préoccupation relevée dans la littérature fait état de la gestion optimale des informations en vue d'aboutir à une meilleure prise de décision [2, 18, 71]. L'utilisation d'informations supplémentaires dans le système est la technique la plus utilisée à cet effet. Cela consiste à utiliser plusieurs objets/capteurs appelées sources physiques de même nature ou encore, utiliser des sources virtuelles (applications, services, etc.).

Nous modélisons ces deux (02) possibilités par la formulation suivante. Considérons un ensemble S de m sources d'information avec $S = \{S_1, S_2..., S_m\}$. Chaque source d'information $S_j \in S$ fournit des informations a une application IoT concernant un ensemble H de n grandeurs physiques (par exemple température, pression, humidité, etc.). Ce fonctionnement est décrit sous la forme matricielle suivante :

$$\begin{array}{c} \\ S_1 \\ \vdots \\ S_j \\ \vdots \\ S_m \end{array} \begin{array}{c} \begin{array}{ccccc} H_1 & \dots & H_i & \dots & H_n \end{array} \\ \begin{pmatrix} M_1^1 & \dots & M_i^1 & \dots & M_n^1 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ M_1^j & \dots & M_i^j & \dots & M_n^j \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ M_1^m & \dots & M_i^m & \dots & M_n^m \end{pmatrix} \end{array} \tag{5.1}$$

où chaque M_i^j désigne l'information de type H_i mesurée par la source d'information S_j. Pour résoudre ce problème, la prise en compte des éléments suivants serait nécessaire :

— les incertitudes dans les informations;

— la fiabilité des sources d'information;

— les informations contradictoires.

Nous avons vu dans les chapitres précédents que dans l'application de la théorie des croyances de Dempster-Shafer permettrait de résoudre les problèmes de qualité d'information susmentionnés. Après avoir proposé une approche pour la quantification de l'incertitude, nous nous intéressons à la résolution de conflit pour assurer la véracité

des informations. Dans ce processus de fusion, la fiabilité des sources d'information est un aspect important dans le processus de fusion. La section suivante présente notre approche pour la résolution des conflits.

5.3 Notre approche de résolution de conflit (ABC_IoT)

Comme mentionné dans la section 5.2, nous devons proposer une solution au problème de la gestion des informations contradictoires issues de plusieurs sources d'information dans un système IoT. Ce problème nous ramenant à un problème de fusion d'information, nous avons mené une étude sur les différentes théories utilisées dans les systèmes distribués comme l'IoT. Comme nous avons vu dans la revue de littérature, la théorie des croyances de Dempster-Shafer, fait preuve de robustesse dans les problèmes de fusion d'information dans les systèmes distribués. Cependant, la résolution du problème de conflit constitue une des issues ouvertes sujette à de nombreuses réflexions. Dans le principe de la théorie des croyances, le conflit entre deux sources d'information peut être défini par la contradiction contenue dans leur réponse exprimée sous la forme de deux degrés de croyance.

Pour résoudre tous ces problèmes, nous proposons une approche dénommée ABC_IoT qui élimine tout problème de conflit entre plusieurs sources contradictoires. L'approche ABC_IoT est présentée sous la forme d'un logigramme (figure 5.1) dont nous détaillerons les parties essentielles.

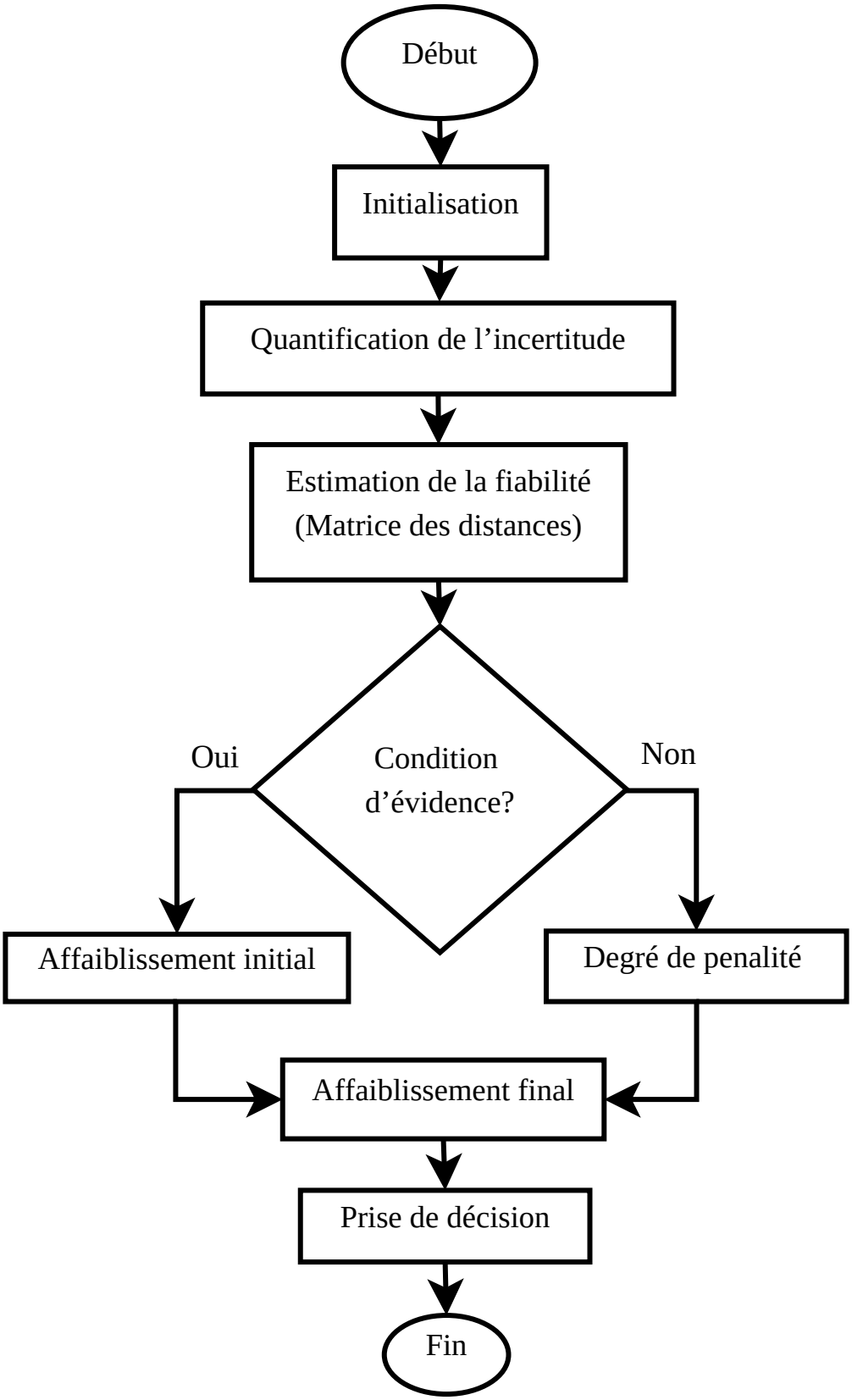

FIGURE 5.1 – Logigramme de l'*ABC_IoT*

Notre approche se base sur deux éléments clés : la modélisation de l'incertitude et de la fiabilité. Concernant la modélisation de l'incertitude, notre méthode utilise la nouvelle mesure de l'incertitude proposée au chapitre 4. Pour quantifier la fiabilité des sources d'informations nous utilisons le principe de fiabilité relative et de fiabilité globale. Le résultat de cette opération aide avec plus de précision, le système de fusion à vérifier et à contrôler les informations provenant des objets et permet de conduire le système à une prise de décision efficace.

5.3.1 Initialisation

L'initialisation consiste à déterminer les éléments de preuve (*BOEs*) selon les informations collectées à partir des sources d'information. La détermination des *BOEs* de chaque source constitue une base de données évidentielle c'est à dire une base dans laquelle se trouvent tous les degrés de croyance correspondant aux informations issues des sources pour la vérification. Pour déterminer les degrés de croyance de chaque source nous utilisons la méthode d'initialisation "IDM".

Exemple 5.1

On considère trois (03) sources S_1, S_2, S_3 qui donnent respectivement les valeurs 21, 5 et 20 degrés celsius qui représentent pour chaque source la température dans la commune de Cocody (Abidjan). En appliquant la méthode "IDM" on obtient les degrés de croyance suivants : $m_{s1} = 0.4468$, $m_{s2} = 0.1064$, $m_{s3} = 0.4255$ $m_{s1,s2,s3} = 0.0213$.

Après avoir déterminé les valeurs d'initialisation, nous passons à la phase de quantification de l'incertitude.

5.3.2 Quantification de l'incertitude

Nous avons vu dans la première contribution que plusieurs facteurs liés à l'incertitude (données bruitées, erronées, manquantes, ...) peuvent entraver la prise de décision. De ce fait, il est crucial de déterminer, dans la démarche de la résolution des conflits, le degré d'incertitude de chaque source d'information. Dans cette étape, nous utilisons la nouvelle mesure de l'incertitude E_{Nm} que nous avons proposée au chapitre 4. Afin d'obtenir un meilleur coefficient d'affaiblissement, la mesure de chaque

fonction de masse est normalisée à partir de l'équation 5.2.

$$\overline{E_{Nm}}(m_i) = \frac{E_{Nm}(m_i)}{\sum_{j=1}^{n} E_{Nm}(m_j)}, j = 1, 2, ..., N \tag{5.2}$$

5.3.3 Estimation de la fiabilité

L'estimation de la fiabilité permet de vérifier la véracité des informations issues de chaque source. L'étude de la fiabilité lorsque l'on dispose de plusieurs sources susceptibles de fournir des informations contradictoires est très importante avant l'étape de la combinaison. Pour ce faire, après la détermination des degrés de croyance, nous utilisons la notion de distance pour estimer la fiabilité de l'information. Ainsi, à partir de la distance de Jousselme [94], nous obtenons une matrice composée des différentes distances calculées comme le présente le modèle suivant:

$$FD_{ij} = \begin{pmatrix} 0 & D_{12} & \dots & D_{1N} \\ D_{21} & 0 & \dots & D_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ D_{N1} & D_{N2} & \dots & 0 \end{pmatrix} \tag{5.3}$$

Où chaque D_{ij} désigne la distance entre la fonction de masse m_i et la fonction de masse m_j.

Fiabilité relative

Dans notre démarche d'estimation de la fiabilité, nous formulons que la fiabilité relative entre deux sources S_1 et S_2 n'est autre que la distance qui sépare leurs fonctions de masse m_1 et m_2. La détermination de la fiabilité relative permet d'observer le comportement d'une source d'information relativement à chacune des autres sources, à partir des informations fournies et représentées par les degrés de croyance. Elle se détermine à l'aide de l'équation 5.4. Cette distance détermine la valeur moyenne sur chaque ligne de la matrice de distance.

$$RF_i = \frac{\sum_{j=1, j\neq i}^{N} D_{ij}}{N-1}, i, j \in 1, 2, ..., N \tag{5.4}$$

Fiabilité Globale

Lorsqu'il y a plusieurs sources, la mesure de la fiabilité correspond à la distance entre la fonction de masse fournie par une source donnée et les autres fonctions de masse. Cette fiabilité consiste en quelque sorte à la fiabilité absolue des sources. La détermination de la fiabilité globale Equation (5.5) permet de déterminer la distance absolue ou globale des différentes sources.

$$GF = \frac{\sum_{i=1}^{N} RF_i}{N} \tag{5.5}$$

5.3.4 Condition d'évidence

La condition d'évidence consiste à déterminer le rapport entre la fiabilité relative RF_i calculée et la fiabilité globale GF. Ainsi, lorsque le rapport $\frac{RF_i}{GF} \leq 1$ alors la condition d'évidence est respectée. Dans le cas contraire, un coefficient de pénalité est déterminé.

5.3.5 Affaiblissement initial

Cette étape constitue l'étape initiale de détermination d'un degré de confiance ou coefficient d'affaiblissement. La détermination du coefficient d'affaiblissement initial dépend de la condition d'évidence. Lorsque la condition d'évidence est respectée, un poids initial est déterminé (Équation 5.6).

$$\delta_i = \exp(-\overline{E_{Nm}}(m_i)) \tag{5.6}$$

où $\overline{E_{Nm}}(m_i))$ est la fonction d'entropie normalisée (c'est à dire, $0 \leq \overline{E_{Nm}}(m_i)) \leq 1$) et m_i désigne la fonction de masse assignée à la source i

Lorsque la condition d'évidence n'est pas respectée, le degré de pénalité s'exprime comme suit :

$$\delta_i = \exp[-(\overline{E_{Nm}^{max}} + 1 - \overline{E_{Nm}}(m_i))] \tag{5.7}$$

où $\overline{E_{Nm}^{max}}$ est l'entropie maximale normalisée.

5.3.6 Affaiblissement final

Cette étape marque la détermination d'un coefficient final d'affaiblissement. Ce poids final est déterminé à partir de l'équation 5.8.

$$\gamma_i = \frac{\delta_i}{\sum_{i=1}^{N} \delta_i}, i = 1, 2, ..., n \tag{5.8}$$

5.3.7 Prise de décision

Cette étape définit le caractère d'aide à la décision à travers la combinaison proprement dite des masses affaiblies (Equation 5.9) à partir de la combinaison conjonctive de Dempster-Shafer (Equation 2.13).

$$\begin{cases} m = \gamma_i m_i \\ \sum_{i=1} \gamma_i = 1 \end{cases} \tag{5.9}$$

5.4 Expérimentations

5.4.1 Cas d'un système intelligent de détection de cibles

L'objectif dans ce cas d'application, consiste à montrer l'efficacité de l'ABC_IoT par rapport à des méthodes bien connues et issues de la théorie de Dempster-Shafer. L'exemple est le suivant :

Exemple 5.2

Dans un système IoT de reconnaissance de cibles, il existe au total trois types de cible $\Omega = \{A, B, C\}$. Supposons que l'on dispose de cinq (5) capteurs. Après avoir récupéré les informations de chaque capteur, on obtient les cinq éléments de preuve BOE suivant :

$m_1 : m_1(A) = 0.41, m_1(B) = 0.29, m_1(C) = 0.30;$

$m_2 : m_2(A) = 0.00, m_2(B) = 0.90, m_2(C) = 0.10;$

$m_3 : m_3(A) = 0.58, m_3(B) = 0.07, m_3(A, C) = 0.35;$

$m_4 : m_4(A) = 0.55, m_4(B) = 0.10, m_4(A, C) = 0.35;$

$m_5 : m_5(A) = 0.60, m_5(B) = 0.10, m_5(A, C) = 0.30.$

Ainsi, à partir des différents degrés de croyance associés à chaque source, le système veut prendre une décision (c'est à dire déduire des trois (03) cibles, laquelle désigne véritablement la cible observée). Après avoir implémenté l'ABC_IoT dans ce cas d'exemple d'application intelligente, nous l'avons comparé à différentes méthodes issues du modèle de Dempster-Shafer notamment les méthodes de résolution de conflits de Dempster et d'autres auteurs bien connus de la littérature [95–97].

Etape 1 : Estimation de la fiabilité à partir de la matrice des distances

Nous déterminons dans un premier temps, la matrice des distances à partir de la distance de Jousselme [94].

$$FD_{ij} = \begin{pmatrix} 0 & 0.5386 & 0.2892 & 0.2699 & 0.2848 \\ 0.5386 & 0 & 0.7195 & 0.6901 & 0.7106 \\ 0.2892 & 0.7195 & 0 & 0.0300 & 0.0255 \\ 0.2699 & 0.6901 & 0.0300 & 0 & 0.0354 \\ 0.2848 & 0.7106 & 0.0255 & 0.0354 & 0 \end{pmatrix}$$

Pour déterminer par exemple, FD_{12} ou FD_{21} à partir de la distance de Jousselme, on a :

$FD_{12} = FD_{21} = d(m1, m2) =$

$$\sqrt{1/2 \begin{pmatrix} 0.41 & 1.19 & 0.4 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 1/2 & 1/2 & 0 & 1/3 \\ 0 & 1 & 1/2 & 0 & 0 & 1/2 & 1/3 \\ 0 & 0 & 1 & 1/2 & 0 & 1/2 & 1/3 \\ 1/2 & 1/2 & 0 & 1/3 & 1 & 1/3 & 2/3 \\ 1/2 & 0 & 1/2 & 1 & 1/3 & 1/3 & 2/3 \\ 0 & 1/2 & 1/2 & 1/3 & 1/3 & 1 & 2/3 \\ 1/3 & 1/3 & 1/3 & 2/3 & 2/3 & 2/3 & 1 \end{pmatrix} \begin{pmatrix} 0.41 \\ 1.19 \\ 0.4 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}}$$

$FD_{12} = FD_{21} = d(m1, m2) = 0.5386$

Après avoir déterminé la matrice des distances, à partir des équations 5.4 et 5.5, nous estimons respectivement les fiabilités relative et globale. Les distances relatives calculées et normalisées donnent : $\overline{d}_1 = 0.3456$, $\overline{d}_2 = 0.6647$, $\overline{d}_3 = 0.2661$, $\overline{d}_4 = 0.2564$,

$\overline{d}_5 = 0.2641$.

Ensuite nous déterminons la distance globale ou la fiabilité absolue des différentes sources d'information $d = 0.3594$.
Nous pouvons remarquer ici, qu'après normalisation, le capteur ayant la fonction de masse m_2 a une distance relative supérieure à la distance globale ($\overline{d}_2 > d$). La source m_2 est donc une source non crédible.

Etape 2 : Quantification de l'incertitude

La nouvelle mesure de l'incertitude que nous avons proposée au chapitre 4, est utilisée pour quantifier l'incertitude dans les éléments de preuve des fonctions m_1, m_2, m_3, m_4 et m_5. On obtient :
$E_{Nm}(m_1) = 1.5664, \quad E_{Nm}(m_2) = 0.469, \quad E_{Nm}(m_3) = 1.7461,$
$E_{Nm}(m_4) = 1.8283, \quad E_{Nm}(m_5) = 1.7168.$
Après normalisation, nous obtenons les résultats suivants :
$\overline{E_{Nm}}(m_1) = 0.2138, \quad \overline{E_{Nm}}(m_2) = 0.0640, \quad \overline{E_{Nm}}(m_3) = 0.2383,$
$\overline{E_{Nm}}(m_4) = 0.2495, \quad \overline{E_{Nm}}(m_5) = 0.2343.$

Etape 3 : Affaiblissement initial

Les sources représentant les fonctions m_1, m_3, m_4 et m_5 étant crédibles, nous utilisons l'équation 5.6 pour les poids initiaux. Les différents calculs donnent :
$\delta_1 = exp(-0.2138) = 0.8075$
$\delta_3 = exp(-0.2383) = 0.7880$
$\delta_4 = exp(-0.2495) = 0.7792$
$\delta_5 = exp(-0.2343) = 0.7911$
La source 2, étant une source non crédible, le calcul de son poids initial donne :
$\delta_2 = exp(-(0.2495 + 1 - 0.0640)) = 0.3056$

Etape 4 : Affaiblissement final

Nous déterminons dans cette étape, les poids finaux que présente chaque fonction de masse. Ainsi, à partir de l'équation 5.8 nous obtenons après normalisation :
$\gamma_1 = 0.2326$
$\gamma_2 = 0.0880$

$\gamma_3 = 0.2270$

$\gamma_4 = 0.2245$

$\gamma_5 = 0.2279$

Comme nous pouvons le constater à partir des valeurs des poids finaux, il y a un conflit entre les éléments de preuve de la source ayant pour fonction de m_2 et les autres sources d'information.

Etape 5 : Prise de décision

La prise de décision dans ce cas d'application consiste à déduire des trois (03) cibles, laquelle représente réellement la cible détectée. Après avoir obtenu les poids finaux, nous déterminons les masses affaiblies à l'aide de l'équation 5.9. Nous obtenons :

$m(A) = 0.4872$, $m(B) = 0.2078$, $m(C) = 0.0790$, $m(AC) = 0.226$.

Enfin, nous déterminons à partir de la règle de combinaison conjonctive de Dempster, les masses finales. Le nombre de sources crédibles étant au nombre de quatre (4), nous combinons les différents éléments de preuve. Le résultat final est :

$m(A) = 0.9878$, $m(B) = 0.0006$, $m(C) = 0.0101$, $m(AC) = 0.0015$.

Après avoir déterminé les degrés de croyance, nous comparons nos résultats à des méthodes bien connues de la littérature. La première méthode est la combinaison de Dempster. Comme nous l'avons noté dans la revue de littérature, c'est la méthode de combinaison de base de la théorie des croyances. La deuxième méthode, la méthode de Murphy *et al.*, est réputée par son approche basée sur une moyenne arithmétique pour la combinaison. La troisième méthode, le méthode de Jiang *et al.*, est une méthode de prétraitement qui se base sur la combinaison de Dempster pour l'étape de combinaison. La particularité de l'ABC_IoT, bien qu'étant une méthode de prétraitement, mesure l'incertitude à travers l'entropie des croyances et aussi met l'accent sur la fiabilité des sources d'information avant l'étape de combinaison.

Les résultats des différentes expérimentations sont présentés dans le tableau 5.1. Dans ce tableau, pour mieux distinguer la précision des méthodes, nous avons fait varier le nombre de sources de deux (02) à cinq (05). La première ligne du tableau donne les degrés de croyance pour deux sources. Dans ce cas, avec les mêmes degrés de croyance, la cible détectée pour les quatre méthodes est B. Lorsque le nombre de sources passe

TABLE 5.1 – Comparaison avec des méthodes utilisant la théorie de Dempster-Shafer

BOE	Méthode	m(A)	m(B)	m(C)	m(AC)	Décision
m_1, m_2	Dempster	0	0.8969	0.1031	0	B
	Murphy *et al.*	0.0964	0.8119	0.0917	0	B
	Jiang *et al.*	0.0964	0.8119	0.0917	0	B
	ABC_IoT	0.0964	0.8119	0.0917	0	B
m_1, m_2, m_3	Dempster	0	0.6575	0.3425	0	B
	Murphy *et al.*	0.4939	0.4190	0.0792	0.0090	A
	Jiang *et al.*	0.7614	0.1295	0.0961	0.0130	A
	ABC_IoT	0.7614	0.1295	0.0961	0.0130	A
m_1, m_2, m_3, m_4	Dempster	0	0.3321	0.6679	0	C
	Murphy *et al.*	0.8362	0.1147	0.0410	0.0081	A
	Jiang *et al.*	0.9379	0.0173	0.0361	0.0087	A
	ABC_IoT	0.9391	0.0168	0.0355	0.0086	A
m_1, m_2, m_3, m_4, m_5	Dempster	0	0.1422	0.8578	0	C
	Murphy *et al.*	0.9620	0.0210	0.0138	0.0032	A
	Jiang *et al.*	0.9837	0.0021	0.0110	0.0032	A
	ABC_IoT	0.9878	0.0006	0.0101	0.0015	A

à trois (03), seule la combinaison de Dempster, conclue que la cible réellement détectée est B. Les autres méthodes, y compris l'ABC_IoT désigne la cible A. Aux quatrième et cinquième lignes du tableau, on obtient respectivement quatre (04) et cinq (05) sources fusionnées. On peut toujours remarquer que les méthodes de Murphy *et al.*, Jiang *et al.* et l'ABC_IoT considèrent que la cible A comme cible détectée par conséquent, l'ABC_IoT donne avec plus de précision, la croyance que la cible est A.

Les figures 5.2 à 5.5 représentent respectivement les degrés de croyance finaux obtenus à partir du tableau 5.1 concernant les différentes cibles observées par les cinq sources d'information. La figure 5.2 représente en fonction du nombre de sources variant de 2 à 5 et de la méthode utilisée, les degrés de croyance déterminés pour la cible A. Pour chacune des méthodes, hormis la méthode de combinaison classique de Dempster, la

cible détectée est A. Cependant comme nous l'avons signifié, la méthode ABC_IoT a un degré de croyance supérieur aux autres degrés. Ce qui fait, la particularité de l'ABC_IoT est que plus le nombre de sources augmente, plus précise est le degré de croyance.

FIGURE 5.2 – Résultat de $m(A)$ basé sur différentes règles de combinaison.

La figure 5.3 représente en fonction du nombre de sources variant de deux (02) à cinq (05) et de la méthode utilisée, les degrés de croyance déterminés pour la cible B. Avec les valeurs observées dans la figure 5.3, hormis la méthode de la combinaison de Dempster, la cible B obtient le plus faible degré de croyance. Les figures 5.4 et 5.5 représentent respectivement les degrés de croyance pour les cibles C et pour les cibles A ou C (c'est à dire $\{A, C\}$). De plus, parmi ces méthodes existantes que nous comparons, la convergence de l'ABC_IoT est la meilleure. Cela est principalement dû à l'utilisation de la fonction de distance pour la fiabilité et la fonction d'entropie améliorée. Ainsi, avec l'ABC_IoT, l'effet des preuves crédibles est extrêmement renforcé, et la " mauvaise " preuve a moins d'influence sur le jugement final.

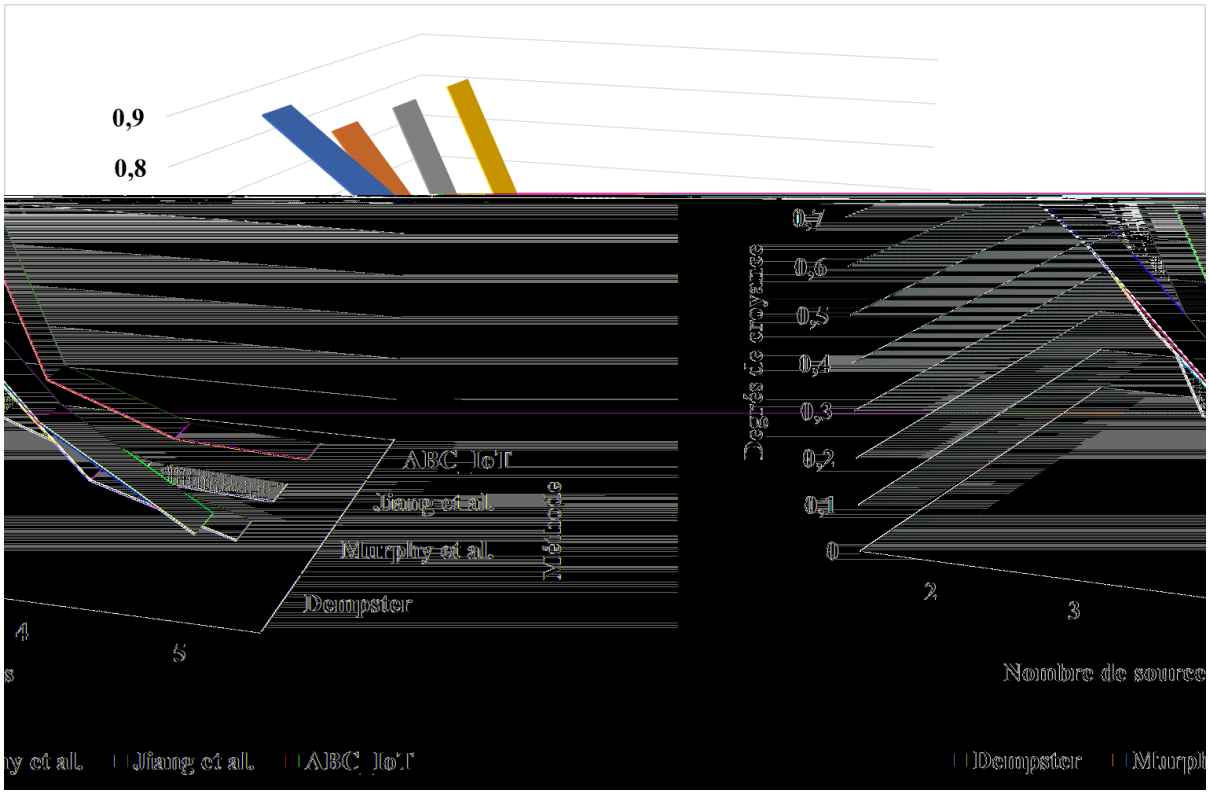

FIGURE 5.3 – Résultat de $m(B)$ basé sur différentes règles de combinaison.

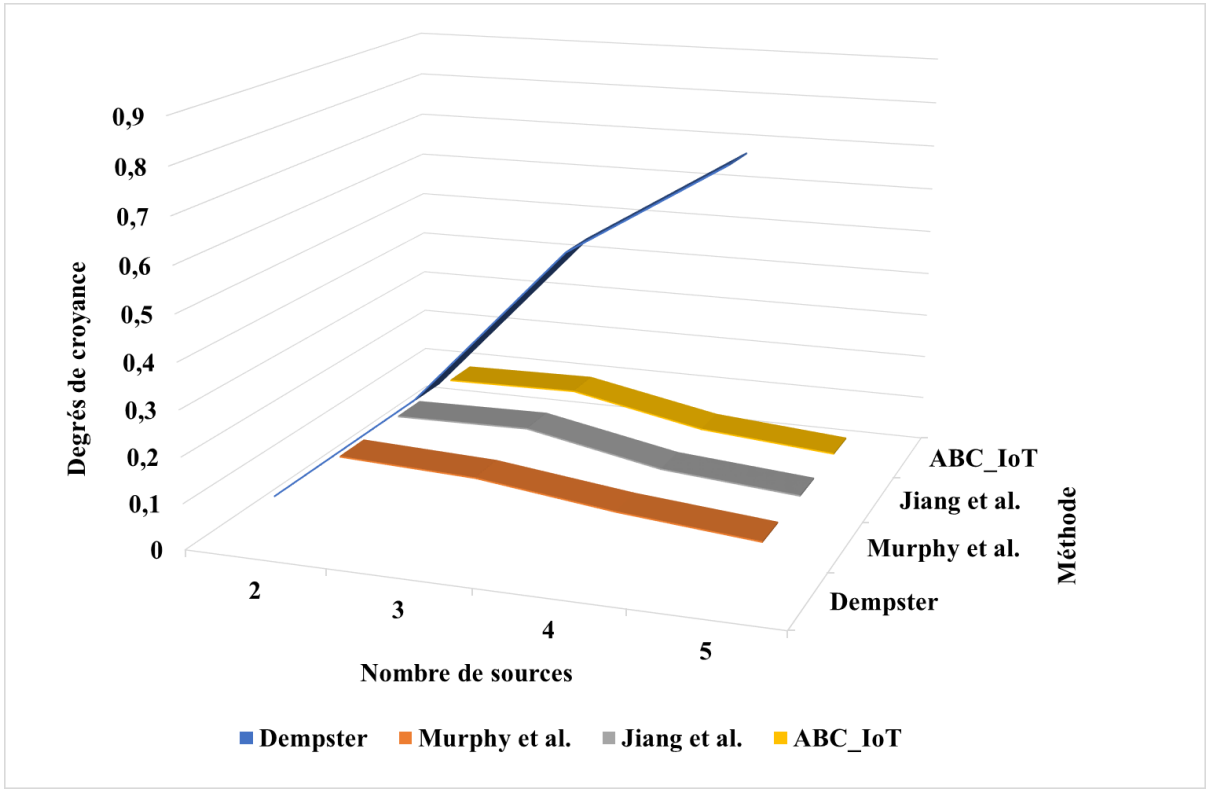

FIGURE 5.4 – Résultat de $m(C)$ basé sur différentes règles de combinaison.

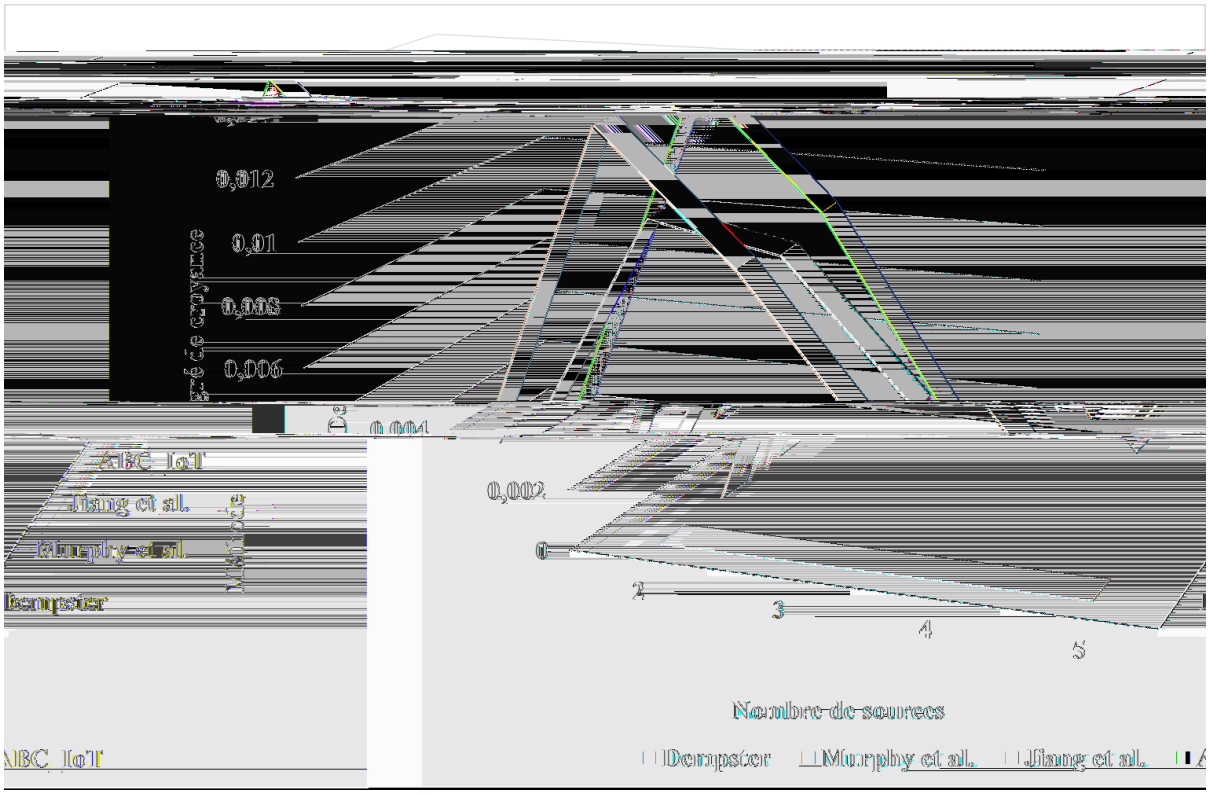

FIGURE 5.5 – Résultat de $m(AC)$ basé sur différentes règles de combinaison.

Enfin, notons que lorsque les degrés de croyance ou éléments de preuve sont très conflictuels, la règle de combinaison classique de Dempster [57] produit des résultats contre-intuitifs qui ne reflètent pas la vérité comme c'est le cas avec la source 2. Et donc il est inutile de l'utiliser dans ce cas d'application qu'après une étape de prétaitrement. Il est aussi important de noter que la moyenne de Murphy *et al.* [82] et la méthode de Jiang *et al.* [80] donnent des résultats relativement raisonnables. Cependant, dans ce cas d'application, notre méthode présente les meilleurs degrés de croyance. Le but étant d'améliorer les méthodes de la théorie des croyances, l'*ABC_IoT* résout efficacement les problèmes de conflit entre plusieurs sources.

5.4.2 Cas d'un système intelligent dans le domaine des Villes Intelligentes

Mise en œuvre

Le deuxième cas d'application est illustré dans le domaine des villes ou communautés intelligentes. L'objectif ici, est de montrer l'efficacité de l'*ABC_IoT* lors du pas-

sage à l'échelle. Il s'agit de montrer la performance de l'ABC_IoT dans des situations où l'application IoT devra fusionner les informations issues de plusieurs de sources d'information disponibles afin de vérifier la véracité des informations issues des objets dont elle a le contrôle direct. Pour l'implémentation de ce scénario, nous utilisons un générateur de données synthétique. Ce générateur de données est une interface de programmation applicative (API) proposée et developpée par Waguih *et al.* [71]. Cette API permet de simuler une grande variété de scénarios où plusieurs sources présentent des comportements différents en termes de taux d'erreur, de niveau de fiabilité, d'informations contradictoires, etc. L'un des éléments importants de ce générateur de données et qu'il garantit un jeu de données vérité-terrain pour permettre l'évaluation en terme de précision des algorithmes de recherche mis en comparaison. Le générateur de données peut produire des données réelles concernant divers domaines d'applications. L'on note par example, le domaine des villes intelligentes avec le jeu de données "Weather" à partir duquel plusieurs sources fournissent des informations sur le temps qu'il fait dans un lieu précis. Pour résoudre le problème de conflits, nous implémentons l'ABC_IoT ainsi que quelques méthodes de découverte de vérité notamment TruthFinder, 2-Estimates et l'ECRM.

Résultats et discussion

Pour montrer la performance de l'ABC_IoT, nous comparons en terme de précision, les résultats après fusion. Le tableau 5.2 présente, en terme de précision, les quatre (04) méthodes, TruthFinder, 2-Estimates, l'ECRM et l'ABC_IoT. Comme signifié dans la description du jeu de données synthétiques, la génération des informations issues des sources s'accompagne d'un jeu de données représentant le modèle vérité-terrain. Ainsi, l'estimation de la précision pour chaque source i est exprimée par $T_{Precision} = \frac{VP_i}{VP_i+FP_i}$, où VP_i désigne une information avérée dans le modèle vérité-terrain et FP_i désigne une information qui se révèle fausse dans le modèle vérité-terrain. La figure 5.6 représente les résultats des méthodes de découverte de vérité les plus connues ainsi que notre méthode de résolution de conflits, sur l'ensemble des données générées. Comme nous pouvons le constater, l'ABC_IoT surmonte en terme de précision, les autres méthodes. On peut également observer que, plus le nombre de sources pour la tâche de fusion augmente, le taux de précision de l'ABC_IoT performe la tâche de fusion. Cependant, ce comportement n'est pas observé avec les autres mé-

TABLE 5.2 – Différentes valeurs de précision

Nombre de sources	TruthFinder	2-Estimates	ECRM	ABC_IoT
20	0.52	0.53	0.8	0.86
40	0.55	0.56	0.82	0.88
60	0.59	0.6	0.85	0.89
80	0.65	0.67	0.86	0.9
100	0.71	0.72	0.88	0.91
120	0.72	0.73	0.9	0.94
140	0.68	0.69	0.9	0.95

thodes où leur performance à tendance à se dégrader lorsque le nombre de sources augmente surtout lorsque le nombre de sources augmente.

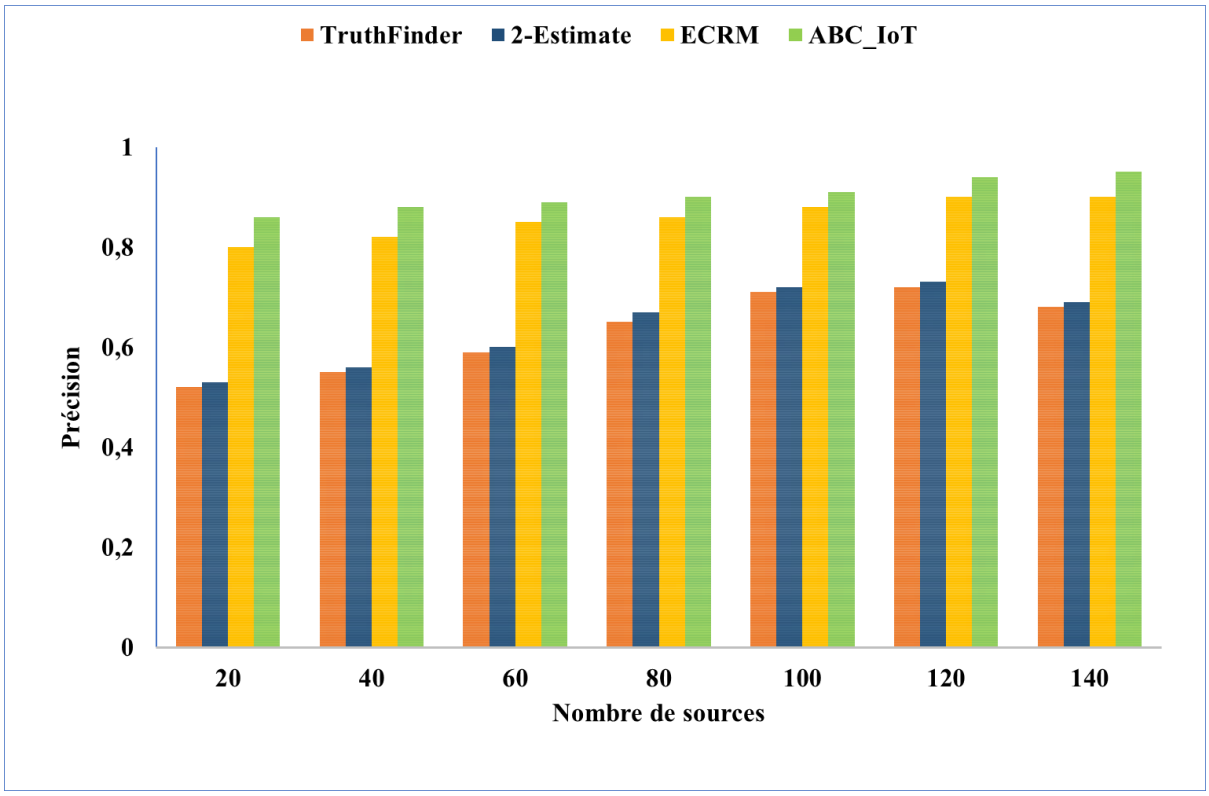

FIGURE 5.6 – Précision de l'*ABC_IoT*

5.5 Conclusion

Avec le développement de l'information moderne et d'Internet, une grande quantité de données et d'informations peut être mise à profit par de multiples sources d'information. Dans les applications IoT, les techniques de fusion d'informations sont de plus en plus importantes pour résoudre le problème lié à la qualité de l'information. Cependant, comment s'assurer de la véracité des informations lorsque plusieurs sources d'information est un problème essentiel pour assurer une meilleur prise de décision. Nous avons, pour ce faire, proposé une méthode de prétraitement dénommée ABC_IoT basée sur la théorie de Dempster-Shafer, une théorie robuste pour la fusion des information en environnement distribué. Avec l'ABC_IoT, la quantification de l'incertitude et la quantification de la fiabilité ont permis de résoudre le problème de conflit. Après expérimentions, l'ABC_IoT présente de meilleurs résultats aussi bien, par rapport aux méthodes issues du framework de Dempster-Shafer qu'à des méthodes récentes de découverte de vérité.

6
Application

6.1 Introduction 86
6.2 Présentation de OpenIoT 87
6.3 Mise en œuvre et résultats 88
6.4 Conclusion 95

6.1 Introduction

Dans ce chapitre, nous illustrons nos contributions dans le domaine des villes intelligentes. Dans ce cas d'application nous nous intéressons particulièrement à quatre (04) types d'information notamment la température, l'humidité, les précipitations et le vent. Pour la mise en place de ce scénario, nous utilisons un intergiciel libre dénommé OpenIoT [98]. Ce chapitre est organisé comme suit. La section 6.2 présente l'outil uti-

lisé pour la mise en application. La section 6.3 décrit la mise en œuvre proprement dite ainsi que les résultats. La section 6.4 conclut ce chapitre.

6.2 Présentation de OpenIoT

OpenIoT est une plateforme fournissant les moyens de formuler et de gérer des environnements comprenant des ressources IoT physiques ou virtuelles [98]. Cette plateforme comprend sept (07) modules principaux tels que présentés à la figure 6.1.

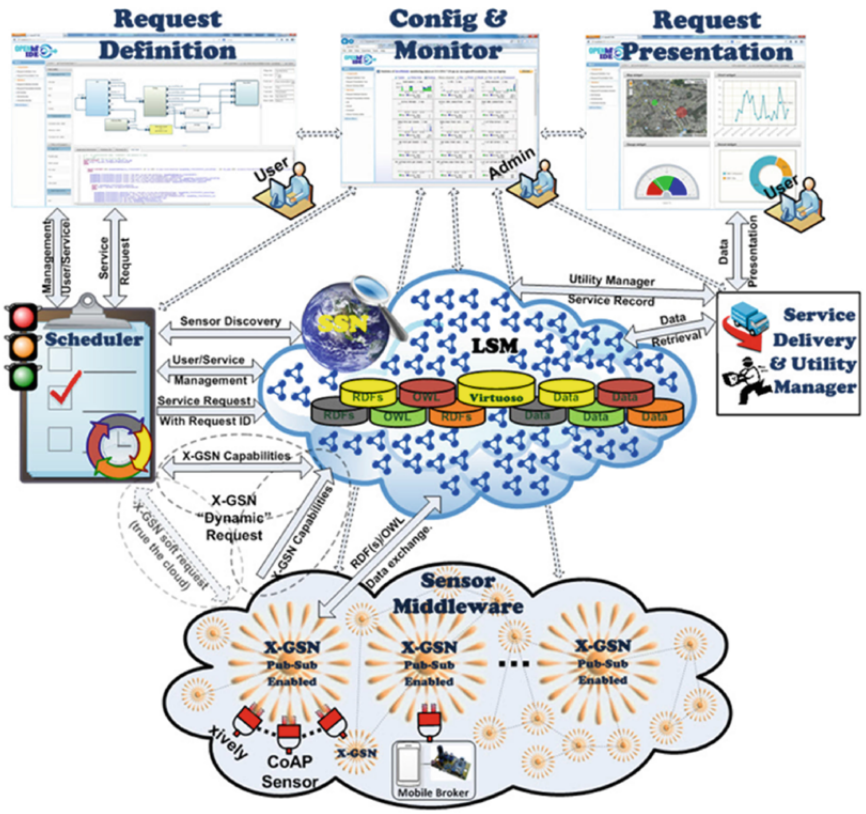

FIGURE 6.1 – Architecture de OpenIoT [98]

— **Sensor Middleware:** Ce module comprend l'intergiciel "Extended Global Sensor Networks, X-GSN)". Il est chargé de la collecte, du filtrage et de la combinaison des flux de données provenant de capteurs virtuels ou des dispositifs physiques;

— **Cloud Data Storage:** Ce module fonctionne comme une base de données dans un Cloud et permet le stockage des flux de données issus des capteurs. L'infrastructure Cloud fournit également des métadonnées pour l'opération du middleware;

— **Scheduler:** Il traite les demandes de déploiement à la demande de services et assure leur bon accès aux ressources dont ils ont besoin. Il découvre les capteurs et les flux de données associés à un service. Il gère également un service et active les ressources impliquées dans sa fourniture;

— **Service Delivery and Utility Manager:** Il combine les flux de données au sein du système OpenIoT afin de livrer le service demandé(Il est généralement exprimé en requête SPARQL);

— **Request Definition:** Ce composant permet la spécification du service demandé à la plateforme OpenIoT. Il comprend un ensemble de services pour spécifier les demandes et les soumet au Planificateur. Ce composant est pris en charge par une interface graphique;

— **Request Presentation:** Ce composant est chargé de la visualisation des sorties d'un service. Il sélectionne les outils nécessaires à partir d'une bibliothèque appropriée afin de faciliter la présentation du service;

— **Configuration and Monitoring:** Ce composant est chargé de la configuration et de la surveillance. Il permet une gestion visuelle des fonctionnalités sur les capteurs et les services qui sont déployés dans la plateforme OpenIoT.

6.3 Mise en œuvre et résultats

6.3.1 Mise en œuvre

Dans OpenIoT, nous créons et configurons quatre (04) capteurs virtuels: un (01) capteur de Température, un (01) capteur d'Humidité, un (01) capteur pour la vitesse du vent et un (01) capteur de Précipitations dans la commune de Cocody (Abidjan - Côte d'Ivoire). Ce test s'est déroulé le 25 Juin 2019 de 10 heures 20 minutes à 10 heures 45 minutes. L'objectif étant d'aider le système à vérifier en temps réel, les informations issues des différents capteurs, nous utilisons le service de recherche de sources "Scheduler". Ce service, se comportant comme une API de recherche et d'extraction de sources d'information web et mobile identifie dix-huit (18) sources ayant des connaissances sur les quatre (04) types d'information météorologiques. Après avoir modifié le comportement du capteur virtuel de température, nous obtenons les informations présentées dans le tableau 6.1.

TABLE 6.1 – Mesures des capteurs

Capteur	10h20-10h25	10h25-10h30	10h30-10h35	10h35-10h40	10h40-10h45
Temp. (°C)	22	16	23	10	26
Hum. (%)	87	87	87	86	86
Vent (Km/h)	23	23	24	24	23
Préc. (%)	23	24	24	23	23

Les tableaux 6.2, 6.3 et 6.4 présentent les informations récoltées par dix-huit (18) sources (des applications web et mobiles). Ces sources dans cette étude constituent simplement des objets dont les informations permettent de vérifier les informations des capteurs. La fusion des informations est effectuée en utilisant l'ABC_IoT.

TABLE 6.2 – Informations issues des sources (1)

Sources	10h20-10h25	10h25-10h30	10h30-10h35	10h35-10h40	10h40-10h45
weather	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
accuweather	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
previmeteo	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
weatherav	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
meteovista	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
meteocity	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24

TABLE 6.3 – Informations issues des sources (2)

Sources	10h20-10h25	10h25-10h30	10h30-10h35	10h35-10h40	10h40-10h45
freemeteo	Temp. (°C) : _ Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : _ Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : _ Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
lachainemeteo	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
worldmeteo	Temp. (°C) : _ Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : _ Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
tameteo	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : _ Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : _ Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
euronews	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
meteoblue	Temp. (°C) : _ Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 23 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24

TABLE 6.4 – Informations issues des sources (3)

Sources	10h20-10h25	10h25-10h30	10h30-10h35	10h35-10h40	10h40-10h45
meteorama	Temp. (°C) : _ Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : _ Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
infoclimat	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
meteocentrale	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
meteocast	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
sat24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24
meteociel	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24	Temp. (°C) : 24 Hum. (%) : 87 Vent (Km/h) : 23 Préc. (%) : 24

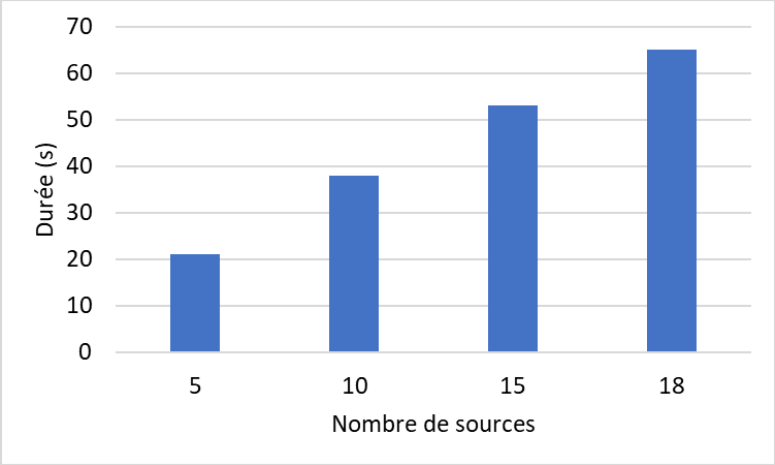

FIGURE 6.2 – Durée de recherche et d'extraction des informations issues des sources

Dans ce travail, nous jugeons la durée de recherche et d'extraction des informations importante pour permettre une deuxième alternative lorsqu'un objet présente des défaillances. Après avoir recherché et extrait les informations, nous passons à l'étape fusion et prise de décision. Dans notre cas, seules les informations de température ont été testées avec le module de fusion intégré dans le module "Service Management". Un autre élément important, c'est la durée afin d'aboutir au résultat final pour la prise de décision. Ainsi, après avoir implémenté l'ABC_IoT dans le "Service Management", nous avons relevé les durées de traitement en fonction du nombres de sources, comme le présente la figure 6.3. Après fusion, les valeurs de température retenues sont mentionnées dans le tableau 6.5.

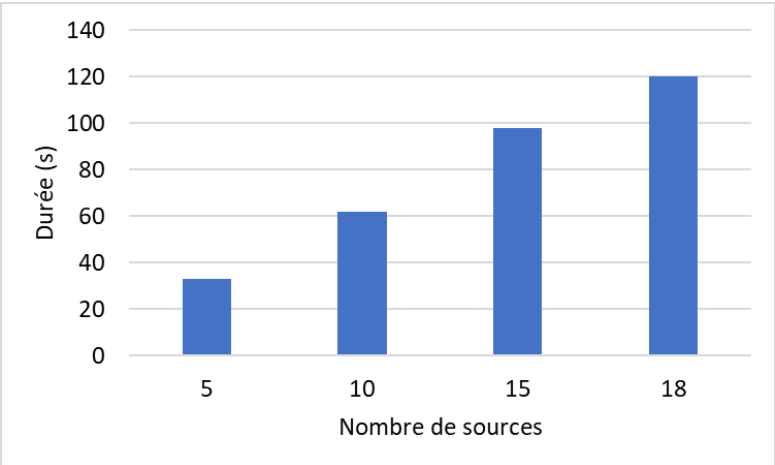

FIGURE 6.3 – Durée de traitement

TABLE 6.5 – Mesures de température retenues après fusion

Capteur	10h20-10h25	10h25-10h30	10h30-10h35	10h35-10h40	10h40-10h45
Temp. (°C)	23	23	23	24	24

La Figure 6.4 présente le graphique représentant les informations issues des capteurs et les informations des sources traitées et fusionnées par intervalle de cinq (05) minutes, 10 heures 20 minutes à 10 heures 45 minutes. Grâce aux informations issues des sources, une bonne décision sans interruption de service peut être prise. Ces expériences ont été réalisées sur un ordinateur portable ayant les caractéristiques suivantes : Pc Acer core i3 ayant 2.4 Ghz de fréquence, 4 Gigaoctets de mémoire RAM et 500 Gigaoctets de Disque Dur.

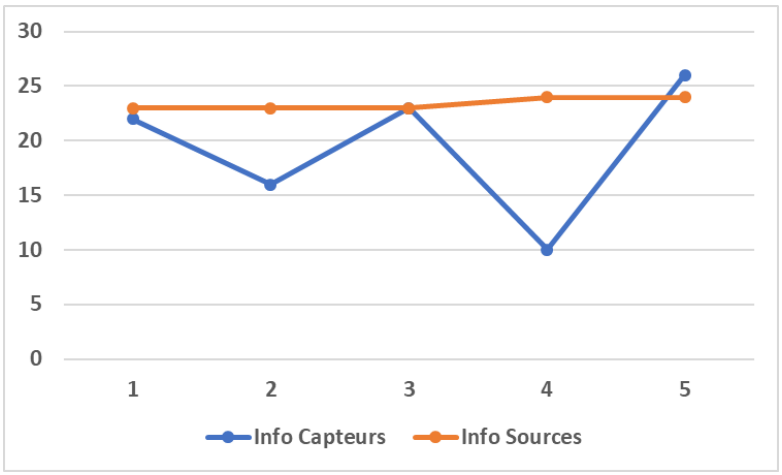

FIGURE 6.4 – Résultat pour la prise de décision

6.4 Conclusion

L'objectif de cette application est de montrer que dans les applications *IoT*, les problèmes d'imprécision dans les informations issues des objets physiques peuvent être résolus et qu'il est possible d'implémenter notre approche dans des applications *IoT* existantes. L'illustration de nos différentes approches suit l'hypothèse selon laquelle, le système *IoT*, via Internet, peut obtenir les informations issues de plusieurs autres sources ou périphériques afin d'améliorer les prises de décision. Ainsi, dans

un scénario des villes intelligentes, avec l'intergiciel OpenIoT, nous avons implémenté l'ABC_IoT. Cette solution permet d'éviter les mauvaises prises de décision.

Conclusion générale

Bilan

Ce manuscrit a été consacré à la proposition d'un modèle en vue d'apporter des solutions aux problèmes liés à la qualité de l'information dans l'Internet des Objets. Dans la littérature, l'incertitude, la fiabilité et la véracité des informations font partie des principaux facteurs qui dégradent la qualité de l'information dans les systèmes distribués. Par ailleurs, plusieurs études ont apporté des solutions à ces problèmes en utilisant les théories de fusion d'information. Parmi ces théories, la théorie de Dempster-Shafer appelée aussi théorie des croyances ou théorie des preuves ou encore théorie de l'évidence, a récemment fait l'objet de nombreuses études liées au traitement de l'information. Cependant dans l'utilisation de celle-ci, de nombreux problèmes demeurent ouverts. Parmi ceux-ci, l'on note le problème de la quantification de l'incertitude et celui de la résolution des conflits.

La première contribution de cette thèse (Chapitre 4), a abordé le problème de la quantification de l'incertitude dans l'utilisation de la théorie de Dempster-Shafer. Après avoir présenté les limites des mesures d'incertitude les plus récentes à savoir les mesures proposées par Deng et Zhou *et al.*, nous avons proposé une nouvelle mesure de l'incertitude. Cette nouvelle mesure, comme les mesures existantes, utilise le principe de l'entropie de Shannon pour la quantification de l'incertitude.

La deuxième contribution de cette thèse (Chapitre 5), a abordé le problème de la véracité des informations dans l'Internet des Objets en utilisant la théorie de Dempster-Shafer. Le principe de cette méthode pour la véracité des informations multisources, consiste à utiliser des informations issues d'autres sources pour améliorer la prise de décision. L'approche basée sur l'entropie des croyances pour la résolution de conflits

dans l'Internet des objets (ABC_IoT) proposée est une méthode de pré-traitement qui utilise deux éléments nécessaires pour une bonne qualité des informations : la quantification des incertitudes des sources d'information et la fiabilité des sources d'information. Concernant la quantification de l'incertitude, nous avons utilisé la nouvelle mesure de l'incertitude proposée au chapitre 4. Concernant la fiabilité, deux notions pour évaluer la confiance au niveau des sources d'information ont été utilisées : la fiabilité relative et la fiabilité globale. Après comparaison de l'ABC_IoT avec les différentes méthodes de découverte de véracité, l'ABC_IoT a montré de meilleures performances.
Nous avons illustré nos contributions dans le domaine des villes intelligentes. L'objectif de cette application est de montrer que l'ABC_IoT peut être implémentée dans les applications IoT. Pour ce faire, nous avons entrepris un scénario de quatre (04) capteurs et dix-huit (18) sources générées dans OpenIoT, un intergiciel libre. L'utilisation de l'ABC_IoT a rendu efficace la prise de décision.

Perspectives

Nous avons réalisé des contributions pour le problème de la qualité des informations en environnement IoT toutefois des défis à savoir, la notion d'apprentissage dans la quantification de l'incertitude et de la fiabilité ainsi que la sécurité, sont à relever.

Apprentissage dans la quantification de l'incertitude, la fiabilité

Dans l'analyse de l'état de l'art, nous avons vu que plusieurs approches permettaient de modéliser l'incertitude dans le modèle de Dempster-Shafer. Afin d'envisager un système bien plus autonome, il serait nécessaire d'établir des modèles en combinant les techniques d'apprentissage à la théorie de Dempster-Shafer. Avec les nouvelles techniques d'apprentissage telle que le Deep learning, les applications IoT pourraient apprendre de leurs erreurs aussi bien dans la quantification de l'incertitude que dans l'estimation de la fiabilité des sources d'information à intégrer dans le système de gestion des informations sensorielles.

Sécurité

Dans l'application de nos modèles, nous n'avons pris en compte qu'une infirme partie de l'aspect de la sécurité à savoir, comment reconnaitre qu'une source est fiable ou encore comment faire confiance à une source d'information. Dans ce cas d'application nous nous sommes uniquement basés sur l'aspect de la véracité des informations. L'application de ce cette technique dégage évidemment une issue ouverte : celle des sources malveillantes. Un aspect sécuritaire plus poussé doit soutenir notre approche pour bonne gestion des applications IoT.

Liste des publications

Publications Scientifiques

Moïse D. MAMBE, Georges N. ANOH, Souleymane OUMTANAGA. *A belief entropy-based approach for conflict resolution in IoT applications*, IEEE International Conference of Smart Cities and Communities, Juillet 2018.

Moïse D. MAMBE, Tchimou N'TAKPE, Georges N. ANOH, Souleymane OUMTANAGA, *A new uncertainty mesure in belief framework*, International Journal of Advance Computer Science and Applications, Octobre 2018.

Congrès Scientifiques

Moïse D. MAMBE, Souleymane OUMTANAGA, Tra Bi GOORE, *Gestion dynamique des données contextuelles dans l'Internet des Objets : Application aux systèmes éducatifs*, Conférence WACREN, Mars 2017.

Moïse D. MAMBE, Souleymane OUMTANAGA, *Une approche dynamique de résolution de conflits pour la fusion de données dans l'IoT*, Olympiades du CAMES, Décembre 2017.

Bibliographie

[1] Aditya Gaur, Bryan Scotney, Gerard Parr, and Sally McClean. Smart city architecture and its applications based on iot. *Procedia computer science*, 52:1089–1094, 2015. *3 citations pages 1, 6, et 7*

[2] M. Ravi, Y. Demazeau, and F. Ramparany. Reasoning with Trust and Uncertainty Illustration in the Internet of Things - IEEE Conference Publication, 2016. *9 citations pages 1, 7, 13, 17, 18, 21, 34, 68, et 69*

[3] W. Cherifi and B. Szafranski. An unsupervised evidential conflict resolution method for data fusion in iot. *Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 819—-823, 2017. *3 citations pages 1, 11, et 68*

[4] Hatem Ben Sta. Quality and the efficiency of data in "Smart-Cities". *Future Generation Computer Systems*, 74:409–416, 2017. *3 citations pages 1, 16, et 55*

[5] F. García, F. Jimenez, and J. Anaya. Distributed pedestrian detection alerts based on data fusion with accurate localization. *Sensors 2013*, pages 11687—-11708, 2013. *cité page 1*

[6] Yuriy V. Shkvarko, Juan I. Yañez, and Gustavo D. Martín del Campo-Becerra. *A Novel Neural Network Computing Based Way to Sensor and Method Fusion in Harsh Operational Environments*. IEEE, 2014. *3 citations pages 1, 18, et 32*

[7] T. Banerjee and D.Swagatam. Multi-sensor data fusion using support vector machine for motor fault detection. *Information Sciences*, 217:96–107, 2015. *2 citations pages 1 et 32*

[8] Arnaud Martin. La fusion d'informations. *Polycopié de cours ENSIETA*, Janvier 2005. *9 citations pages 2, 19, 24, 29, 30, 31, 40, 41, et 42*

[9] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7):1645–1660, 2013. *3 citations pages 6, 9, et 10*

[10] K. Ashton et al. That 'internet of things' thing. *RFID journal*, 22(7):97–114, 2010. *cité page 6*

[11] N. Kara, Djoudi Touazi, et al. *Conception d'un réseau de communication pour une maison intelligente en utilisant la technique d'internet des objets.* PhD thesis, Université A/Mira de Bejaia, 2017. *2 citations pages 6 et 7*

[12] Herve Costantini, Eric Renault, and Selma Boumerdassi. Future internet in the clouds. In *2011 11th Annual International Conference on New Technologies of Distributed Systems*, pages 1–5. IEEE, 2011. *cité page 8*

[13] Chen and Yen-Kuang. *Challenges and Opportunities of Internet of Things*. IEEE, 2012. *cité page 8*

[14] Hugues Sansen. Le big bang d'un écosystème protéiforme: L'e-santé. *Springer-Verlag Berlin Heidelberg*, 2014. *cité page 9*

[15] Maciej Kranz. *Building the internet of things: Implement new business models, disrupt competitors, transform your industry*. John Wiley & Sons, 2016. *cité page 9*

[16] M. S. Fox and C. J. Pettit. On the completeness of open city data for measuring city indicators. *Smart Cities Conference (ISC2),IEEE First International*, 18(2):1–6, 2015. *3 citations pages 11, 15, et 16*

[17] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The Internet of Things: A survey. *Computer Networks*, 54(15):2787–2805, 2010. *cité page 13*

[18] A. Gyrard, K. Datta, C. Bonnet, and K. Boudaoud. *Cross-Domain Internet of Things Application Development: M3 Framework and Evaluation*. IEEE, 2015. *3 citations pages 14, 15, et 69*

[19] B. Cai, Y. Liu, and Y. Zhang. Application of bayesian networks in quantitative risk assessment of subsea blowout preventer operations. *Risk Anal.*, 33:1293–1311, 2016. *3 citations pages 15, 28, et 52*

[20] A. Caragliu, C. Del Bo, and P. Nijkamp. Smart cities in europe. *Journal of urban technology*, 18(2):65–82, 2015. *cité page 16*

[21] A. Ojo, Z. Dzhusupova, and E. Curry. Exploring the nature of the smart cities research landscape. *Smarter as the New Urban Agenda, Springer*, pages 23–47, 2016. *cité page 16*

[22] M. Kehoe, M. Cosgrove, S. Gennaro, C. Harrison, W. Harthoorn, and J. Hogan. Smarter cities series: A foundation for understanding ibm smarter cities. *Redguides for Business Leaders, IBM*, 2015. *cité page 16*

[23] A. Ojo, E. Curry, and T. Janowski. Designing next generation smart city initiatives. *harnessing findings and lessons from a study of ten smart city programs*, 2014. *cité page 16*

[24] S. Munir and J. Stankovic. A survey on context-aware systems. *DepSys: Dependency Aware Integration of Systems for Smart Homes*, 2014. *2 citations pages 17 et 27*

[25] John A. Stankovic. Research directions for the internet of things. *IEEE Internet of Things Journal*, 1(1):3–9, 2014. *2 citations pages 17 et 18*

[26] W. Xu, W. Trappe, Y. Zhang, and T. Wood. The feasibility of launching and detecting jamming attacks in wireless networks. *Proc. of MobiHoc*, pages 46–57, 2012. *2 citations pages 17 et 27*

[27] A. Toumi. Intégration des bases de connaissances dans les systèmes d'aide à la décision : Application à l'aide à la reconnaissance de cibles radar noncoopératives. *PhD thesis, Université de Bretagne Occidentale*, 2017. *cité page 18*

[28] W. Jiang, Y. Yang, Y. Luo, and X. Qin. Determining basic probability assignment based on the improved similarity measures of generalized fuzzy numbers. *Int. J. Comput. Commun. Control*, 10:333–347, 2015. *cité page 18*

[29] G. Rogova and E. Bossé. Information quality in information fusion. *13th Conference on Information Fusion (FUSION)*, pages 1–8, 2010. *2 citations pages 18 et 49*

[30] A. Martin and C. Osswald. Toward a combination rule to deal with partial conflict and specificity in belief functions theory. *In International Conference on Information Fusion*, 2007. *5 citations pages 19, 21, 23, 30, et 42*

[31] Willliam Feller. *An introduction to probability theory and its applications*, volume 2. John Wiley & Sons, 2008. *2 citations pages 19 et 55*

[32] A. Rahman and S. Hailes. A distributed trust model. *Workshop on New Security Paradigms*, pages 48–60, 2016. *cité page 19*

[33] C. Walid and S. Bolesław. An unsupervised evidential conflict resolution method for data fusion in iot. *Proceedings of the Federated Conference on Computer Science and Information Systems pp*, pages 819—823, 2017. *4 citations pages 20, 34, 43, et 50*

[34] Shilpa Gite and Himanshu Agrawal. *On Context Awareness for Multisensor Data Fusion in IoT*. Springer, 2016. *4 citations pages 21, 30, 31, et 32*

[35] Otman Basir and Xiaohong Yuan. Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory. *Information Fusion*, 8(4):379–386, 2014. *3 citations pages 24, 55, et 56*

[36] George J. Klir. Generalized information theory. *Fuzzy Sets and Systems*, 40(1):127–142, 2010. *2 citations pages 26 et 55*

[37] Deng and Yong. D numbers: theory and applications. *Journal of Information &Computational Science*, 9(9):2421–2428, 2014. *5 citations pages 26, 30, 43, 55, et 68*

[38] A. Galland, S. Abiteboul, A. Marian, and P. Senellart. Corroborating information from disagreeing views. *Proceedings of WSDM'10, ACM*, pages 131—-140, 2014. *3 citations pages 26, 51, et 68*

[39] X. Qin and Y. Gu. Data fusion in the internet of things. *Procedia Eng*, 15:3023—-3026, 2014. *cité page 27*

[40] A. Josang. The right type of trust for distributed systems. *Workshop on new security paradigms, ACM*, pages 119–131, 1996. *cité page 27*

[41] M. Georgeff, B. Pell, M.Pollack, and M. Tambe. Intelligent agents v springer-verlag lecture notes in ai. *In Intelligent Agents V: Agents Theories, Architectures, Springer*, 1998. *cité page 29*

[42] M. Georgeff and F. Ingrand. Real-time reasoning: The monitoring and control of spacecraft systems. *IEEE Int Conf. of Artificial Intelligence Applications*, pages 198–204, 1990. *cité page 29*

[43] D. Bellot. *Fusion de Données Avec Des Réseaux Bayésiens Pour La Modélisation Des Systèmes Dynamiques et Son Application En Télémédecine*. PhD thesis, Université Henri Poincaré-Nancy I, 2002. *2 citations pages 29 et 42*

[44] M. Mambe, S. Oumtanaga, and G. Anoh. A belief entropy-based approach for conflict resolution in iot applications. In *2018 1st International Conference on Smart Cities and Communities (SCCIC)*, pages 1–5. IEEE, 2018. *cité page 29*

[45] T. Redman. Data quality : Management and technology. *Bantam Books*, 1992. *cité page 29*

[46] M. Chebbah, A. Martin, and B. Yaghlane. About sources dependence in the theory of belief functions. *In the 2nd International Conference on Belief Functions (BELIEF'2012)*, 2012. *cité page 29*

[47] H. Liu and Q. Lin. Fuzzy failure mode and effects analysis using fuzzy evidential reasoning and belief rule-based methodology. *IEEE Trans. Reliab.*, 62:23—-36, 2013. *cité page 29*

[48] Deyun Zhou, Yongchuan Tang, and Wen Jiang. A modified belief entropy in Dempster-Shafer framework. *PLOS ONE*, 12(5):e0176832, 2017. *11 citations pages 29, 43, 47, 55, 56, 57, 59, 60, 61, 64, et 66*

[49] Claude E Shannon. A note on the concept of entropy. *Bell System Tech. J*, 27(3):379–423, 1948. *4 citations pages 31, 43, 45, et 55*

[50] R. Kalman and others. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960. *cité page 32*

[51] R. Kalman and R. Bucy. New results in linear filtering and prediction theory. *Journal of basic engineering*, 83(1):95–108, 1961. *cité page 32*

[52] C. Harris and Q. Gan. Multisensor data fusion using Kalman filters based on neuro-fuzzy linearisation. In *Multisensor Fusion*, pages 435–456. Springer, 2012. *2 citations pages 32 et 42*

[53] Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. Using Bayesian networks to analyze expression data. *Journal of computational biology*, 7(3-4):601–620, 2014. *cité page 32*

[54] Xiu Wang1 Jian Rong1 Xiaochun Zhong. A Hybrid Neural Network-Fuzzy Logic Architecture for Multisensor Data Fusion in Target Tracking System. *IEEE Journal of information fusion*, 2007. *cité page 32*

[55] John F. Lemmer. Confidence factors, empiricism and the Dempster-Shafer theory of evidence. *arXiv preprint arXiv:1304.3437*, 2013. *cité page 32*

[56] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976. *4 citations pages 35, 40, 50, et 55*

[57] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339, 1967. *4 citations pages 35, 50, 55, et 82*

[58] Didier Dubois and Henri Prade. Properties of measures of information in evidence and possibility theories. *Fuzzy Sets and Systems*, 24(2):161–182, 1987. *4 citations pages 36, 46, 55, et 66*

[59] Kari Sentz and Scott Ferson. *Combination of evidence in Dempster-Shafer theory*. PhD thesis, Binghamton University, 2014. *2 citations pages 36 et 39*

[60] Philippe Smets and Robert Kennes. The transferable belief model. *Artificial intelligence*, 66(2):191–234, 1994. *4 citations pages 40, 47, 50, et 52*

[61] R. Yager. On the dempster-shafer framework and new combination rules. *Inf Sci 41*, pages 93—-137, 1987. *cité page 40*

[62] D. Dubois and H. Prade. *Possibility Theory*. Springer New York, 2014. *2 citations pages 41 et 55*

[63] Abdelmounaim Debieche and Karima Zebouchi. Utilisation de la théorie de dempster et shafer pour la fusion d'information. Mémoire de Master en informatique. USTHB, 2016. *cité page 42*

[64] P. Smets. Data fusion in the transferable belief model. *Université Libre de Bruxelles*, 1:21–33, 2000. *2 citations pages 42 et 50*

[65] Ee-Peng Lim, Jaideep Srivastava, and Shashi Shekhar. An evidential reasoning approach to attribute value conflict resolution in database integration. *IEEE Transactions on Knowledge and Data Engineering*, 8(5):707–723, 1996. *cité page 43*

[66] M. Mambe, T. N'Takpe, N. Anoh, and S. Oumtanaga. A new uncertainty measure in belief entropy framework. *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, 9(11):600–606, 2018. *cité page 43*

[67] Jianping Yang, Hong-Zhong Huang, Qiang Miao, and Rui Sun. A novel information fusion method based on Dempster-Shafer evidence theory for conflict resolution. *Intelligent Data Analysis*, 15(3):399–411, 2015. *2 citations pages 43 et 50*

[68] Yong Deng. Deng entropy. *Chaos, Solitons & Fractals*, 91:549–553, 2017. *11 citations pages 43, 46, 50, 55, 56, 59, 60, 61, 63, 64, et 66*

[69] D. A. Waguih, N. Goel, H. M. Hammady, and L. Berti-Equille. AllegatorTrack: Combining and reporting results of truth discovery from multi-source data. *2015 IEEE 31st International Conference on Data Engineering*, pages 1440–1443, 2015. *3 citations pages 43, 51, et 52*

[70] Thierry Denoeux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172:234–264, 2011. *cité page 43*

[71] L. Berti-Equille. Scaling up truth discovery. *IEEE 32nd International Conference on Data Engineering*, pages 1418—-1419, 2016. *6 citations pages 43, 51, 52, 68, 69, et 83*

[72] U. HOHLE. Entropy with respect to Plausibility Measures. *Proc. of 12th IEEE Int. Symp. on Multiple Valued Logic, Paris, 1982*, 1982. *3 citations pages 46, 55, et 66*

[73] Ronald R. Yager. Entropy and Specificity in a Mathematical Theory of Evidence. *International Journal of General Systems*, 9(4):249–260, 1998. *3 citations pages 46, 55, et 66*

[74] George J. Klir and Arthur Ramer. Uncertainty in the Dempster-Shafer Theory: A Critical Re-examination. *International Journal of General Systems*, 18(2):155–166, 1998. *3 citations pages 46, 55, et 66*

[75] J. Klir and B. Parviz. *A note on the Measure of Discord*. Springer New York, 2010. *3 citations pages 46, 55, et 66*

[76] Thomas George and Nikhil R. Pal. Quantification of Conflict in Dempster-Shafer Framework: A new approach. *International Journal of General Systems*, 24(4):407–423, 1996. *2 citations pages 46 et 66*

[77] M. Smithson. Ignorance and uncertainty : Emerging paradigms. *Springer Verlag*, pages 45–55, 1989. *cité page 48*

[78] P. Bonnisone and R. Tong. reasoning with uncertainty in expert systems. *International Journal of Man-Machine Studies*, pages 241–250, 1985. *2 citations pages 48 et 52*

[79] E. Lefebvre, M. Hadzagic, and E. Bossé. Advances and challenges in multisensor data and information processing. *chapitre On quality of information in multi-source fusion environments,IOS Press*, pages 57–60, 2007. *cité page 49*

[80] Wen Jiang, Miaoyan Zhuang, Xiyun Qin, and Yongchuan Tang. Conflicting evidence combination based on uncertainty measure and distance of evidence. *SpringerPlus*, 5(1):1217, 2016. *2 citations pages 50 et 82*

[81] Lotfi A. Zadeh. Toward a generalized theory of uncertainty (GTU)—an outline. *Information Sciences*, 172(1):1–40, 2013. *2 citations pages 50 et 55*

[82] C. Murphy. Combining belief functions when evidence conflicts. *Decision Support Systems*, 29(1):1–9, 2010. *2 citations pages 50 et 82*

[83] R. Yager. On the dempster-shafer framework and new combination rules, 1987. *cité page 50*

[84] L. Berti-Equille and J. Borge-Holthoefer. Veracity of data: From truth discovery computation algorithms to models of misinformation dynamics. *Morgan and Claypool Publishers*, 2015. *3 citations pages 51, 52, et 68*

[85] X. Yin, J. Han, and S. Philip. Truth discovery with multiple conflicting information providers on the web. *IEEE TKDE*, 20(6):796—-808, 2016. *2 citations pages 51 et 68*

[86] Huadong Wu, Mel Siegel, Rainer Stiefelhagen, and Jie Yang. *Sensor fusion using Dempster-Shafer theory [for context-aware HCI]*, volume 1. IEEE, 2015. *2 citations pages 55 et 56*

[87] Mathias Bauer. Approximation algorithms and decision making in the Dempster-Shafer theory of evidence—An empirical study. *International Journal of Approximate Reasoning*, 17(2-3):217–237, 1997. *2 citations pages 55 et 56*

[88] Dennis M Buede and Paul Girardi. A target identification comparison of Bayesian and Dempster-Shafer multisensor fusion. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27(5):569–577, 1997. *2 citations pages 55 et 56*

[89] Jianwei Wang, Fuyuan Xiao, Xinyang Deng, Liguo Fei, and Yong Deng. Weighted Evidence Combination Based on Distance of Evidence and Entropy Function. *International Journal of Distributed Sensor Networks*, 12(7):3218784, 2016. *cité page 55*

[90] David Harmanec and George J. Klir. Measuring Total Uncertainty in Dempster-Shafer Theory: A Novel Approach. *International Journal of General Systems*, 22(4):405–419, 1994. *cité page 56*

[91] S. Li, L. Xu, and S. Zhao. The internet of things: a survey. *Inf Syst Front*, 17:243–259, 2016. *cité page 68*

[92] Wen Jiang, Jun Zhan, Deyun Zhou, and Xin Li. A method to determine generalized basic probability assignment in the open world. *Mathematical Problems in Engineering*, 2016, 2016. *cité page 68*

[93] M. Chebbah and B. Yaghlane. Reliability estimation based on conflict for evidential database enrichment. *Workshop on the theory of belief functions*, 2012. *cité page 68*

[94] A. Jousselme, D. Grenier, and E. Bosse. A new distance between two bodies of evidence. *Information Fusion*, 2:91–101, 2001. *2 citations pages 73 et 76*

[95] Kaijuan Yuan and Yong Deng. Conflict evidence management in fault diagnosis. *International Journal of Machine Learning and Cybernetics*, 10(1):121–130, 2019. *cité page 76*

[96] Jian Wang, Kuoyuan Qiao, Zhiyong Zhang, and Fei Xiang. A new conflict management method in dempster–shafer theory, 2017. *cité page 76*

[97] Zhang, Limao, Ding, Lieyun, and Wu. An improved Dempster–Shafer approach to construction safety risk perception, 2017. *cité page 76*

[98] J. Soldatos, N. Kefalakis, and M. Hauswirth. Openiot: Open source internet-of-things in the cloud. *Springer International Publishing Switzerland 2015*, pages 13–25, 2015. *2 citations pages 86 et 87*