PEOPLE’S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
BLIDA -1- UNIVERSITY
INSTITUTE OF AERONAUTICS AND SPATIAL STUDIES

Dissertation

Submitted in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy -**DLMD**- in **Aeronautics**
Option: **Avionics**

Presented by **Nour El Imane HAMDA**

Theme

Managing Big IoT Data in Smart Environments

Defended in front of the respected dissertation-committee composed of:

President	Salah BOUKRAA	Professor	Blida-1- University
Thesis director	Allel HADJALI	Professor	ENSMA, France
Thesis director	Mohand LAGHA	Professor	Blida-1- University
Examiner	Abdelkarim MEZIANE	Professor	CERIST
Examiner	Hamid AZZOUNE	Professor	USTHB

Blida, May 21st, 2024

Abstract

Data quality is crucial in IoT-based smart environments, where the reliability and accuracy of information collected from interconnected devices significantly influence system effectiveness and efficiency. Multisensor data fusion technique has emerged as a powerful tool for managing imperfect data from heterogeneous sources, thereby enhancing operational efficiency and enabling effective decision-making. The Dempster–Shafer (DS) theory of evidence provides a robust and flexible mathematical framework for modeling and fusing uncertain, imprecise, and incomplete data. However, Dempster′s combination rule can lead to counterintuitive results when dealing with highly conflicting data sources.

This thesis focuses on enhancing data quality within IoT-based smart environments by investigating data fusion techniques for managing heterogeneous IoT data. It specifically addresses the complexities arising from highly conflicting data sources within the Dempster-Shafer theory framework. Novel solutions are proposed in this work to overcome limitations associated with Dempster′s combination rule and to improve the quality, reliability, and utility of data from multiple sensors. These solutions involve preprocessing the original evidence model by assigning weighting factors to evaluate the reliability of each information source, considering both uncertainty and conflict using various metrics.

To demonstrate the validity and effectiveness of the proposed approaches, simulations are conducted across various domains, including fault diagnosis, IoT decision-making, and situational awareness within UAV systems. Additionally, a comparative analysis with several similar methods from existing literature is carried out to validate the efficiency and superiority of the proposed solutions in terms of conflict management effectiveness, convergence, fusion result reliability and decision accuracy. These findings contribute to achieving more robust and trustworthy outcomes in dealing with complex and conflicting data.

Key words: *Internet of things ; Data quality ; Multisensor data fusion ; Dempster-Shafer theory ; Fuzzy logic ; Conflict management ; Weighted average evidence*

Résumé

La qualité des données est cruciale dans les environnements intelligents basés sur l'internet des objet, où la fiabilité et la précision des informations collectées à partir de dispositifs interconnectés ont une influence significative sur l'efficacité et l'efficience du système. La fusion de données multicapteurs est un outil puissant pour gérer les données imparfaites provenant de sources hétérogènes, ce qui permet d'améliorer les performances opérationnelles et de prendre des décisions efficaces. La théorie de l'évidence de Dempster-Shafer (DS) fournit un cadre mathématique robuste et flexible pour la modélisation et la fusion de données incertaines, imprécises et incomplètes. Cependant, la règle de combinaison de Dempster peut conduire à des résultats contre-intuitifs lorsqu'il s'agit de sources de données hautement conflictuelles.

Cette thèse porte principalement sur l'amélioration de la qualité des données dans les environnements intelligents basés sur l'internet des objets (IoT) à travers l'étude des techniques de fusion de données multicapteurs pour la gestion des données hétérogènes de l'IoT. Elle traite spécifiquement les complexités résultant de sources de données hautement conflictuelles dans le cadre de la théorie de Dempster-Shafer. De nouvelles solutions sont proposées dans ce travail, pour surmonter les limites de la règle de combinaison de Dempster et pour améliorer la qualité, la fiabilité et l'utilité des données provenant de multiples capteurs. Ces solutions consistent à prétraiter le modèle de preuve initial en attribuant des facteurs de pondération pour évaluer la fiabilité de chaque source d'information, en tenant compte à la fois de l'incertitude et du conflit à l'aide de diverses métriques.

Pour démontrer la validité et l'efficacité des approches proposées, elles ont été appliquées dans divers domaines, notamment le diagnostic de pannes, la prise de décision dans l'internet des objets et la conscience situationelle dans les systèmes de drones. En outre, une analyse comparative avec plusieurs méthodes similaires issues de la littérature existante est réalisée pour valider la supériorité des solutions proposées en termes d'efficacité de gestion des conflits, de convergence, de fiabilité des résultats de fusion et de précision des décisions. Cette analyse contribue à obtenir des résultats plus robustes et fiables lors du traitement de données complexes et conflictuelles.

Mots clés: *Internet des objets ; Qualité des données ; Fusion de données multicapteurs ; Théorie de Dempster-Shafer ; Logique floue ; Gestion des conflits ; Preuves moyennes pondérées*

ملخص

تُعد جودة البيانات أمرًا بالغ الأهمية في البيئات الذكية القائمة على إنترنت الأشياء، حيث تؤثر موثوقية ودقة المعلومات التي يتم جمعها من الأجهزة المترابطة بشكل كبير على فعالية النظام وكفاءته. وقد برزت تقنية دمج البيانات من عدة مستشعرات كأداة قوية لإدارة البيانات غير الكاملة من مصادر غير متجانسة، وبالتالي تعزيز الكفاءة التشغيلية وتمكين اتخاذ القرارات الفعالة. توفر نظرية ديمبسترشافر للأدلة إطارًا رياضيًا قويًا ومرنًا لنمذجة ودمج البيانات غير المؤكدة وغير الدقيقة وغير المكتملة. غير أن قاعدة ديمستر للدمج يمكن أن تؤدي إلى نتائج غير منطقية عند التعامل مع مصادر بيانات جد متعارضة.

تتمحور هذه الأطروحة حول تحسين جودة البيانات في البيئات الذكية القائمة على إنترنت الأشياء من خلال دراسة تقنيات دمج البيانات من عدة مستشعرات لإدارة بيانات إنترنت الأشياء غير المتجانسة. وتعالج على وجه التحديد التعقيدات الناشئة عن مصادر البيانات شديدة التعارض في إطار نظرية ديمبسترشافر. وفي هذا العمل، تم اقتراح حلول جديدة للتغلب على القيود المرتبطة بقاعدة ديمبستر للدمج ولتحسين جودة وموثوقية وفائدة البيانات التي يتم جمعها من عدة مستشعرات. تتضمن هذه الحلول، المعالجة المسبقة لنموذج الدليل الأصلي من خلال تعيين أوزان لتقييم موثوقية كل مصدر معلومات، مع الأخذ بعين الاعتبار كلاً من عدم اليقين والتعارض باستخدام مقاييس مختلفة.

ولإثبات صلاحية وفعالية الطرق المقترحة، تم تطبيقها في مجالات مختلفة، بما في ذلك تشخيص الأعطال، واتخاذ القرارات ضمن إنترنت الأشياء، والوعي الظرفي ضمن أنظمة الطائرات بدون طيار. بالإضافة إلى ذلك، تم إجراء تحليل مقارن مع العديد من الطرق المماثلة من الأبحاث السابقة للتحقق من كفاءة الحلول المطروحة وتفوقها من حيث فعالية إدارة التعارض والتقارب وموثوقية نتائج الدمج ودقة القرار. تساهم هذه الاستنتاجات في التوصل إلى نتائج أكثر فعالية وموثوقية في التعامل مع البيانات المعقدة والمتعارضة.

كلمات مفتاحية: إنترنت الأشياء؛ جودة البيانات؛ دمج البيانات من عدة مستشعرات؛ نظرية ديمبسترشافر؛ المنطق الضبابي؛ إدارة التعارض؛ المتوسط المرجح للأدلة.

Acknowledgement

Above all, I thank Allah the Almighty for blessing me with the faith, patience, and strength necessary to pursue my research.

I express my gratitude and dedicate this work to my parents, who have supported me unconditionally throughout this journey. Their unwavering belief in me has surpassed even my own, and without them, none of this would have been achievable.

I would also like to thank my esteemed supervisors; Pr. Allel Hadjali, for his guidance and support, and Pr. Mohand Lagha, for his constant help and motivation, as well as for his exemplary human qualities.

This thesis was conducted under the co-supervision of the Aeronautical Sciences and Spatial Studies Institute at Blida 1 University and LIAS Laboratory at ENSMA School, France. My sincere gratitude goes to the members of the Aeronautical Sciences Laboratory (LSA), headed by Pr. Tahar Rezoug, and to all the members of the institute, under the guidance of Pr Amina Benkhedda, for providing an environment conducive to intellectual innovation. Additionally, I am thankful to all members of the LIAS laboratory at ENSMA school, headed by Pr Emmanuel Grolleau, for their warm welcome and provision of the necessary research facilities.

I am also grateful to the thesis defense committee members, chaired by Pr Salah Boukraa, for accepting to evaluate this work. Their invaluable feedback has significantly enriched my research, enabling me to identify and address the weaknesses.

Special thanks are due to my sisters, Selma and Yasmine, my journey companion, Zahra Ragoub, for their unwavering love, encouragement, and understanding. Their support carried me through the tough times and kept me motivated to strive for excellence.

Lastly, I thank my friends and everyone who has been part of my journey and believed in me. Your support has been instrumental in shaping my academic and personal growth.

Contents

List of Figures

List of Tables

List of Symbols

$\mathbf{p}(\mathbf{H_i}/\mathbf{m_j})$	Conditional probability
$\mathbf{E}$	Space of hypothesis
$\mathbf{H_i}$	Hypothesis
$\mu_A(x)$	Membership function
$\mathbf{\Omega}$	Frame of discernment
$\mathbf{2^{\Omega}}$	Power set
$\mathbf{A_i}$	Subset (Hypothesis)
$\lvert\mathbf{A_i}\rvert$	Cardinality of a subset
$\mathbf{m}(\mathbf{A_i})$	Mass function
$\mathbf{Bel}(\mathbf{A_i})$	Belief function
$\mathbf{Pl}(\mathbf{A_i})$	Plausibility function
$\mathbf{k}$	Conflict coefficient
α_i	Discounting factor
$\mathbf{d_H}$	Hellinger distance
$\mathbf{d_{IH}}$	Improved Hellinger distance
$\mathbf{d_J}$	Jousselme distance
$\mathbf{E_s}$	Shannon entropy
$\mathbf{E_d}$	Deng etropy
$\mathbf{D}$	Jaccard matrix
$\mathbf{Sim}(\mathbf{m_i}, \mathbf{m_j})$	Similarity degree
$\mathbf{Sup}(\mathbf{m_i})$	Support degree
$\mathbf{CRD}(\mathbf{m_i})$	Credibility degree
$\mathbf{I_v}(\mathbf{m_i})$	Information volume
α	Threshold
$\mathbf{N}$	Number of pieces of evidence
$\mathbf{n}$	Number of hypotheses
$\mathbf{D_j}(\mathbf{m_i}, \mathbf{m_j})$	Distance matrix
$\mathbf{cos}(\mathbf{m_i}, \mathbf{m_j})$	cosine value
$\mathbf{w_I}$	Initial weights

$\mathbf{w}$	Final weights
$\mathbf{m_w}(\mathbf{A})$	weighted averaged mass function
BetP	Pignistic probability
$\mathbf{L_s}$	Visual distance
$\boldsymbol{\alpha}_s$	Visual angle
$\boldsymbol{\lambda}$	Uncertainty coefficient of visual distance
$\boldsymbol{\sigma}$	Uncertainty coefficient of visual angle

List of Acronyms

IoT	Internet of Things
ITU	International Telecommunication Union
RFID	Radio frequency identification
DQ	Data quality
ANN	artificial neural networks
JDL	Joint Directors of Laboratories
AI	Artificial Intelligence
DS	Dempster Shafer theory
FOD	Frame of discernment
BOE	Body of evidence
BPA	Basic Probability Assignment
Bel	Belief function
Pl	Plausibility function
Dec	Decision
UAV	Unmanned Aerial Vehicle
OBS	Obstacle

General Introduction

Context and motivation of the research:

A smart environment is a cohesive interconnected small world where various devices, objects, and systems collaborate intelligently to achieve common goals, improving efficiency, user experience, and overall functionality. It encompasses a wide range of applications, including smart homes, smart transport, smart cities... etc, providing advanced services for industrial production and making human lives more comfortable. The Internet of Things (IoT) on the other hand, is a paradigm that connects multiple and diverse smart objects via internet, it has extended the internet's vision by enabling the connection of people and physical objects anytime, anywhere, with anything or anyone, using any path, any network, and any service.

Recently, significant research endeavors have been dedicated to the seamless integration of the Internet of Things (IoT) with smart environments. Interconnected smart devices equipped with embedded sensors, actuators and communication technologies, capable of exchanging and sharing data through the internet are integrated to enable the collection, analysis and utilization of data to optimize processes, services and interactions. This collaborative integration presents a compelling advancement, as it extends the capabilities of smart objects and augments the overall functionality of intelligent ecosystems, enabling the user to monitor the environment from remote sites.

In IoT-based smart environments, tremendous volumes of data are continuously generated by the smart devices using their sensors every single second. Referred to as "big data," these IoT data are characterized by high-volume, high-velocity, high-variety, and high-veracity properties. Due to multiple factors such as environmental noise, sensor defects, or calibration errors, the data gathered are prone to various imperfections, making them noisy, uncertain, conflicting, or even erroneous.

Ensuring data quality is of utmost importance in IoT-based smart environments. The reliability and accuracy of information collected from interconnected devices significantly impact the effectiveness and efficiency of the entire system. Inaccurate or unreliable data can lead to flawed analyses, misguided decision-making, and suboptimal performance of smart applications. Addressing these challenges necessitates advanced technologies for effective data management, enabling the extraction of accurate knowledge and valuable insights.

A myriad of techniques exist for IoT data management and processing. Data fusion plays a vital role in enhancing the quality and usefulness of the processed data,it aims to combine data gathered from various heterogeneous sources in the best possible manner to get more accurate and consistent information. Data fusion is defined as the theory, techniques and tools that are used for combining sensor data, or data derived from sensory data, into a common representational format.

Various mathematical methods are employed in the data fusion process, classified primarily into three categories: probability-based, artificial intelligence-based, and evidence-based techniques [1]. Evidence theory, also known as the theory of belief functions or Dempster-Shafer theory (DS theory), stands out as a robust and flexible mathematical tool for modeling and merging uncertain, imprecise, and incomplete data. Dempster first introduced the theory in 1967 [2] as a generalization of Bayesian inference, later expanded by his student Shafer in 1976 [3] into a comprehensive framework for uncertain reasoning.

DS theory finds extensive application in numerous multisensor data fusion applications,decision making [4, 5, 6], fault diagnosis [7, 8, 9], target recognition [10, 11, 12] . . . etc., owing to its flexibility and effectiveness in handling uncertainty problems and its ability in merging heterogeneous data obtained from multiple sources without prior knowledge using Dempster's combination rule. However, the application of DS evidence theory has its own limitations. Dempster's combination rule generates counterintuitive results when dealing with highly conflicting data.

The problem addressed in this thesis is mainly related to data management in the context of IoT-based smart environments. More specifically, it is about the study of data fusion and mining technologies for integrating the IoT data coming from heterogeneous sources. Based on Dempster Shafer theory, robust and comprehensive evidence combination solutions are proposed to combine multiple bodies of evidence and overcome the problem of coflict encountered by the classical Dempster's combination rule.

Objective of the research:

The main objectives of this work are:

- Design of robust fusion operators for multisensor data fusion in IoT environments that manage data imperfections; reduce uncertainties and handle conflicts among data sources, improve data quality to enhance system performance and ensure reliable and accurate decision making.

- Confirm the effectiveness of the proposed solutions by presenting and discussing the obtained simulation results.

- Provide a rich bibliography to assist the future studies in the domain.

Thesis outline

The manuscript is organized as follows:

Chapter 1 offers a comprehensive overview of key concepts related to the subject: the Internet of Things, data quality, and multisensor data fusion.

Chapter 2 comprises a literature review and a comparative study of the mathematical methods for data fusion, and the selection of the framework that best aligns with IoT applications.

Chapter 3 is devoted to the presentation of the designed improved evidence combination approaches.

Chapter 4 presents the simulation results and comparative analysis.

Finally, the conclusion section summarizes and discusses the research conducted in this thesis, presenting the results obtained. It also outlines potential avenues for future research and areas yet to be explored.

1

General background & state of the art

Contents

1.1 Introduction

In order to better understand the research conducted in this thesis, the following sections provide definitions for all the theoretical concepts used in the study. The chapter introduces the basic concepts of the Internet of Things (IoT), along with the notions that arise from the quality of data generated in IoT-based smart environments. Additionally, an exploration of the tools used for data processing is provided, with a particular emphasis on the basic principles of multisensor data fusion technique.

1.2 Internet of things IoT

Internet of Things (IoT) has become a highly promising and impactful technology in the present era. It has extended the internet's vision by enabling the connection of physical objects in our environment to the Internet. The Cluster of European Research Projects on the Internet of Things (CERP-IoT) defined the IoT as allowing "people and things to be connected Anytime, Anyplace, with Anything and Anyone, ideally using Any network and Any service" [13]. IoT refers to a vast network of interconnected physical devices, objects, and systems that have the ability to autonomously collect, communicate, share data, all while seamlessly interacting with one another through the network without human intervention. This global connectivity has been achieved by embedding sensors, actuators, micro-controllers, and other components within these objects that can range from household appliances and wearable devices to industrial machinery and infrastructure. Connecting these objects and enabling them to gather and share data streamlines and automates complex tasks that may surpass human capabilities. This, in turn, leads to the extraction of valuable insights, automation of processes, and the ability to make more informed decisions. The recent advancements and miniaturization in electronics and computers have played a vital role in making the realization of IoT evident, paving the way for the emergence of a diverse spectrum of innovative applications in various domains, including cities and homes, environmental monitoring, transportation systems, healthcare, and more, leading to substantial savings in time, resources, and effort. Figure 1.1 illustrates some significant applications of IoT.

1.2.1 Brief history of the Internet of things (IoT)

The concept of the Internet of Things (IoT) has evolved over several decades. In the early 1980s, the idea of interconnecting intelligent devices emerged. In 1982, at Carnegie Mellon University in the United States, a connected vending machine was used to check and report beverage inventory [14], representing a notable advancement towards the Internet of Things.

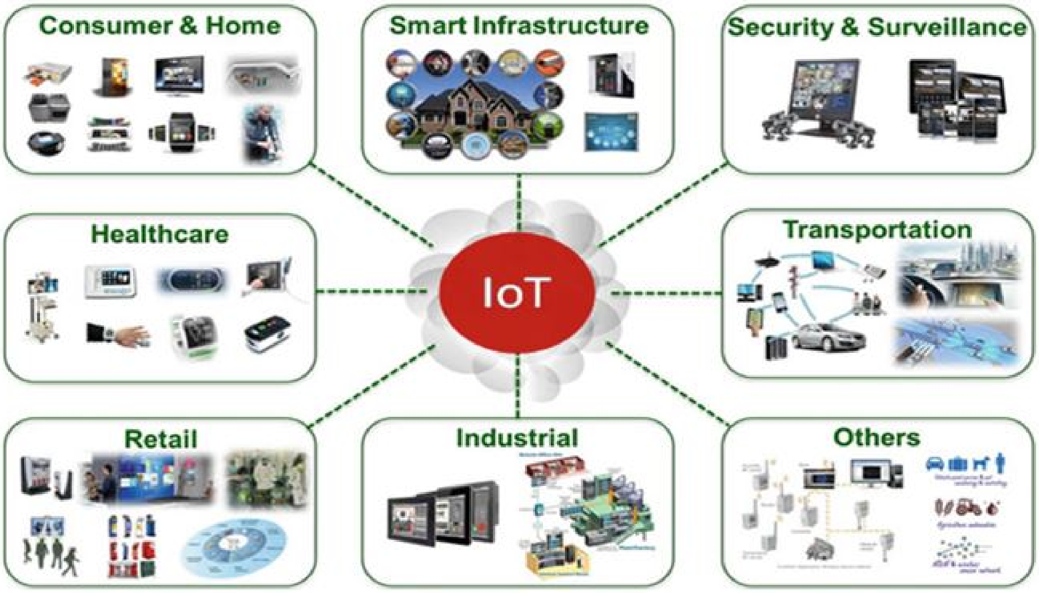


Figure 1.1 – *IoT applications*

In 1991, Mark Weiser, a computer scientist at Xerox PARC, published an article titled "The Computer of the 21st Century," where he introduced the concept of ubiquitous computing. He envisioned an environment where computers would seamlessly integrate into our everyday lives, enabling smooth interaction between users, computers, and the surrounding objects. Although Weiser did not explicitly mention the IoT in his article, he laid the conceptual foundation for the future development of this technology [15].

Later in 1999, Kevin Ashton, co-founder of the Auto-ID Center at the Massachusetts Institute of Technology (MIT), coined the term "Internet of Things" (IoT) [16] in the context of optimizing logistics processes, the main idea he put forward was that real-world objects could be equipped with sensors and radio frequency identification (RFID) chips, enabling them to send and receive data via the Internet. These objects could be identified, tracked, and remotely controlled, opening up a wide range of possibilities for automation and interaction. This idea was a significant milestone in the evolution of the IoT and paved the way for many potential applications.

Throughout the 2000s, technological advancements fueled the growth of the IoT. Wireless networks improved, offering more robust and extensive connectivity. Sensors were miniaturized, becoming smaller and more affordable, enabling their integration into a wide variety of objects. Additionally, the declining production costs of electronic devices made IoT solutions more accessible. In 2005, the International Telecommunication Union (ITU) started to focus on the IoT by proposing a vision of intelligent connectivity and sensing capabilities among real-world objects. This initiative contributed to the recognition and promotion of the IoT as a key area of technological development.

In the last decade, the adoption of the IoT has experienced exponential growth. Industries and governments have started recognizing the potential of the IoT to enhance efficiency, productivity, and quality of life. Sectors such as home automation (smart homes), smart cities, connected healthcare, precision agriculture, manufacturing, and logistics have embraced IoT solutions to optimize their operations and deliver new services.

Today, The IoT has become an ubiquitous reality, with an increasing number of devices and objects connected to the Internet. Smart objects collect data, interact with each other and users, and enable advanced functionalities such as remote control, real-time data analysis, automation, and AI-driven decision-making. The IoT is constantly evolving, with new technological advancements, innovative applications, and challenges to address, opening up new possibilities in many areas of everyday life and industry.

1.2.2 Definitions

The term Internet of Things was firstly coined by Kevin Ashton in 1999 to describe a tracking system that was designed by attaching RFID tags to some objects and connecting them to the internet. He defined it as : " the integration of sensors connected to the Internet, behaving in a similar manner to the Internet, by making open ad hoc connections, freely sharing data, and enabling access to various applications. This allows computers to understand the world around them and become the nervous system of humanity [16]". Ashton's definition emphasizes the connectivity, data sharing, and the ability of IoT devices to interact with the physical world, enabling a deeper understanding and integration of technology into our daily lives. Since then, several definitions have been proposed that reflect different perceptions of the IoT.

In [17], it is defined as a network of networks that enables, through standardized and unified electronic identification systems and wireless mobile devices, the direct and unambiguous identification of digital entities and physical objects. This allows for the retrieval, storage, transfer, and processing of data associated with them without discontinuity between the physical and virtual worlds. In [15] Internet of Things (IoT) is defined as a network that interconnects ordinary physical objects with the identifiable addresses so that provides intelligent services. Atzori et al. [18], identify the IoT paradigm as the result of the convergence of three main visions; the first one represents things oriented perspective which deals with various communication infrastructures connecting devices, systems, and users. The second dimension corresponds to the internet oriented vision, which focuses on enabling interaction between intelligent physical objects and users. Third, the semantic oriented vision which is related to how to represent, store, interconnect and retrieve useful information from massive and inconsistent data generated by IoT. Another definition by Gubbi et al. [19], which focuses on the Internet of Things (IoT) for smart environments as an interconnected system of devices that prioritize the needs and experiences of users is provided as follows: “IoT for smart environments is an interconnection of sensing and actuating devices, providing the ability to share information across platforms through a unified framework, developing a common operating picture for enabling innovative applications. This is achieved by seamless large-scale sensing, data analytics, and information representation using cutting edge ubiquitous sensing and cloud computing". The RFID group define IoT as: The worldwide network of interconnected objects uniquely addressable based on standard communication protocols. and according to the ITU (International Telecommunication Union) [20], the Internet of Things is defined as "a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies." while now, The paradigm of IoT is simplified as any-time, any-place, and any-one connected [21] (see figure 1.2).

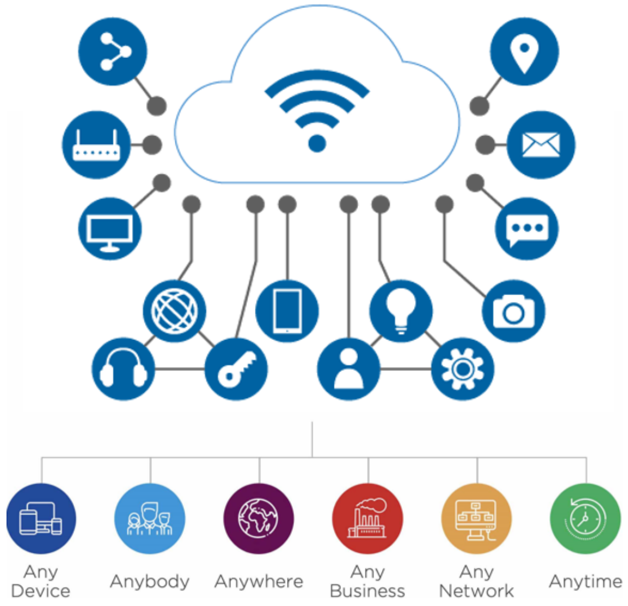


Figure 1.2 – *Illustration of the Definition of the Internet of Things*

Although diverse definitions and explanations exist for understanding IoT, there is no standard, unified, and shared definition of the Internet of Things. Some definitions emphasize the technical aspects of IoT, while others focus more on its uses and functionalities, but all of them commonly agree upon the idea, that the Internet of Things (IoT) represents a diversified paradigm that involves the interaction between a diversity of hardware and software technologies to create new applications/services aiming at accomplishing common goals.

1.2.3 IoT architecture

An IoT (Internet of Things) system typically consists of several key components that work together to enable connectivity and communication between physical devices and the internet. IoT (Internet of Things) architecture refers to the structure or framework that outlines the components, layers, and interactions involved in an IoT system. In literature, there are several common IoT architecture models, where the classification was applied in three-, four-, five-, six- or seven-layer models, but one widely used approach is the three-tier architecture that consists of the Perception, Network/Transmission and Application Layer, as depicted in Figure 1.3, in which the layers cannot be regarded as sub-layers, but rather as distinct components that collectively capture the essential aspects of IoT functionality[22].

1.2.3.1 Perception Layer

Also called device layer, it is the closest layer to the physical world. It consists of sensor-enabled physical objects in an IoT based system, capable of sensing, actuating, communicating, monitoring, and controlling. Sensors gather data regarding location, motion & environmental changes such as temperature, humidity, pressure...etc, from the surrounding. Actuators, on the other hand, enable physical actions based on the received instructions. Gathered data are converted into digital signals and transmitted to the network layer to ensure a secure transfer to the data processing unit.

1.2.3.2 Network Layer

Also known to be the Transmission or connectivity layer. The network layer guarantees for the secure transmission of the data gathered from sensors to the data processing system and enables connectivity and data transfer between sensors, actuators, gateways, and other devices. It encompasses various communication technologies, protocols, and network infrastructure components. . It can include wired connections (Ethernet, Powerline) or wireless connections (Wi-Fi, Bluetooth, Zigbee, cellular networks) depending upon the sensor devices and based on the requirements of the IoT system.

1.2.3.3 Application Layer

The application layer is where data from the perception layer is processed, analyzed, and utilized to derive insights or take actions. This layer typically consists of cloud platforms, software applications, and services that manage and interpret the collected data. Cloud platforms provide scalable storage, computing power, and analytics capabilities for processing large volumes of IoT data. The application layer also includes user interfaces, and integration with external systems to enable interactions and control between users, devices, and services.

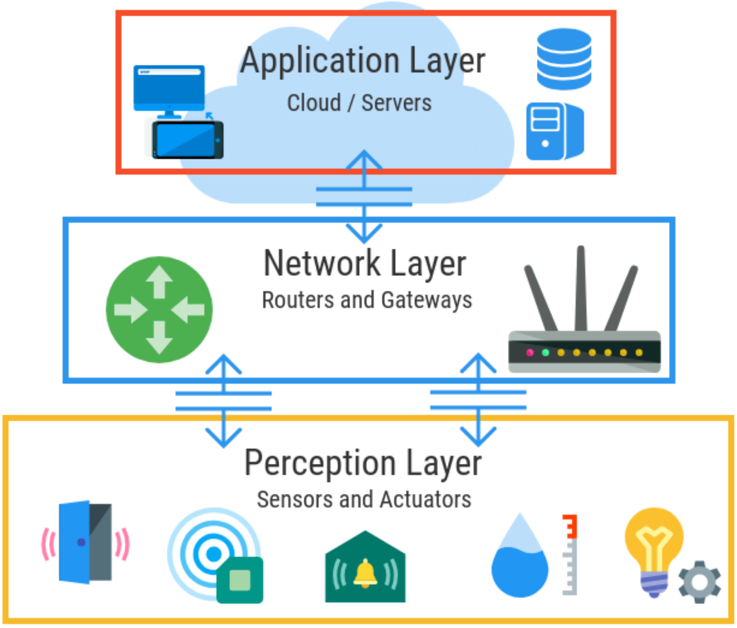


Figure 1.3 – *IoT three-layer architecture [23]*

In addition to the three-tier architecture, there are variations and additional layers that can be included based on the specific use case, scale, and requirements of an IoT system. For instance, some architectures include an edge computing layer, which brings computational capabilities closer to the devices or sensors, reducing latency and improving real-time processing. Security is a cross-cutting concern across all layers of IoT architecture and should be implemented at each level to protect data, devices, and communications.

1.2.4 Challenges and issues

In today's world, a multitude of IoT devices connect through networks to deliver essential information to users. However, implementing IoT solutions is not a simple task, as it brings forth numerous challenges beyond security. In the following, we will provide a concise overview of some of the primary challenges associated with IoT implementations that need to be addressed to ensure efficient and secure operations.

1.2.4.1 Security

IoT devices often have limited computing resources, making them vulnerable to security breaches. Weak authentication, lack of encryption, and inadequate security controls can expose IoT devices and networks to cyberattacks. Protecting data privacy, ensuring device integrity, and preventing unauthorized access are critical challenges.

1.2.4.2 Privacy and Ethical Concerns

IoT environments collect and process vast amounts of personal and sensitive data. Ensuring privacy protection, obtaining user consent, and complying with privacy regulations pose challenges. Addressing ethical concerns related to data usage, consent, transparency, and potential biases is crucial for building trust in IoT deployments.

1.2.4.3 Scalability

IoT networks typically involve a massive number of devices generating large volumes of data. Scaling the infrastructure to handle the increasing number of devices and the associated data traffic can be complex. Managing network congestion, data storage, and processing requirements are significant scalability challenges.

1.2.4.4 Interoperability

Interoperability refers to the capacity of two systems to effectively communicate, exchange information, program, and transfer data between each other, enabling the implementation of the provided data [24]. IoT ecosystems involve a diverse range of devices from different manufacturers, operating on various communication protocols. Ensuring seamless interoperability and integration across devices, platforms, and networks is a major challenge. Standardization efforts are crucial to enable device compatibility and simplify development and deployment.

1.2.4.5 Power Constraints

Many IoT devices are battery-powered or operate on limited power sources. Optimizing power consumption and extending device battery life are critical challenges. Energy transparency between software development and hardware presents a promising solution for tackling power constraints. This transparency is established through the creation of a bridge connecting hardware and software, enabling seamless interoperability. Through this bridge, accurate estimation of energy consumption is ensured, enabling the continuous operation of a device [25]

1.2.4.6 Network Infrastructure

IoT devices rely on robust and reliable network connectivity. However, IoT deployments may face challenges in areas with limited network coverage or inconsistent connectivity. Ensuring sufficient network infrastructure, including wireless protocols, connectivity options, and network reliability, is essential.

1.2.4.7 Data Management and Analytics

IoT environments generate vast amounts of data from sensors, devices, and systems. Effectively managing and analyzing this data to derive actionable insights and valuable information pose challenges. Data processing, storage, real-time analytics, and data integration across diverse sources are important considerations as the overall performance of the application is highly dependent on the quality of the provided data.

1.3 Data quality

Data plays a pivotal role as a valuable asset within various domains, serving as a fundamental source for extracting valuable insights, delivering services, and enabling effective communication. However, it is important to acknowledge that the true value of data lies not only in its abundance but also in its quality. Low data quality levels can significantly impact the overall effectiveness of the associated data applications, leading to flawed analyses and erroneous conclusions.

1.3.1 Definitions

Data quality (DQ) is a multidisciplinary field with the aim of achieving high standards in decisions making and actions taking [26]. While there is no universally accepted definition of data quality, various definitions have been reported in the literature. Notably, Philip Crosby, a prominent quality advocate, defined quality as "conformance to requirements" way back in 1979. Dr. Thomas Redman asserts that data can be considered of high quality when they are fit for their intended purposes in operations, decision-making, and planning, leading to the definition of data quality as "fitness for use" [27].

Furthermore, data quality is described as "the degree to which information has content, form, and time characteristics, which give it value to specific end-users," as mentioned in [28]. In [29], quality is regarded as "the degree to which information is meeting user needs according to external, subjective user perceptions." It is also defined as "meeting or exceeding customer expectations" , or "satisfying the needs and preferences of its users," in [30]. According to the ISO standard, quality refers to the entirety of the attributes or characteristics of an entity that directly impact its capability to meet both stated and implied requirements [31]. For IoT domain, data quality essentially refers to the suitability of the collected data from smart devices for delivering ubiquitous services to IoT users. It is evident that all these definitions are users-centric, these users may be either humans or automated systems.

Data quality encompasses a range of factors, including accuracy, completeness, reliability, consistency, and timeliness... etc, Without ensuring high-quality data, the insights derived from them may be flawed or misleading, potentially resulting in ineffective strategies, inaccurate predictions, or compromised decision-making processes.

1.3.2 Data quality dimensions

Data quality dimensions refer to the various aspects or characteristics of data elements that can be defined, quantified, measured, implemented, and monitored [27]. These dimensions help assess the overall reliability, accuracy, and usefulness of the data. There exist a multitude of data quality dimensions, and the choice of relevant ones varies depending on the application. The commonly recognized data quality dimensions include: accuracy, completeness, consistency, timeliness, validity, reliability, relevance, uniqueness and integrity

1.3.2.1 Accuracy:

Refers to the degree to which data correctly represents the real-world objects or events it is supposed to capture. Accurate data is devoid of errors, inconsistencies, and discrepancies.

1.3.2.2 Completeness:

Reflects the extent to which data captures all the required information without any missing or incomplete values. Complete data provides a comprehensive and holistic view of the subject matter.

1.3.2.3 Consistency

Indicates the coherence and harmony of data across different sources, systems, or time periods. Consistent data ensures that there are no contradictions or conflicts when comparing or integrating data from multiple sources.

1.3.2.4 Timeliness

Reflects the relevance and currency of data in relation to its intended use. Timely data is up-to-date and available in a timely manner, enabling timely decision-making and analysis.

1.3.2.5 Validity

Refers to the degree to which data adheres to defined rules, constraints, or standards. Valid data meets predefined criteria and is suitable for its intended purpose.

1.3.2.6 Reliability

Indicates the trustworthiness and dependability of data. Reliable data is accurate, consistent, and free from biases or errors, providing a high level of confidence in its usability.

1.3.2.7 Relevence

Reflects the alignment between the data and the specific requirements or needs of the intended users. Relevant data is meaningful, applicable, and contributes directly to the desired outcomes or objectives.

1.3.2.8 Uniqueness

Refers to the absence of duplicate or redundant data. Unique data ensures that each piece of information is represented only once, reducing data redundancy and improving efficiency.

1.3.2.9 Integrity

Indicates the overall quality, soundness, and completeness of data. Data integrity ensures that data is protected from unauthorized modifications, deletions, or corruption, maintaining its accuracy and reliability.

1.3.3 IoT Data Characteristics

In IoT environment, a gigantic amount of data with complex structures called 'Big data' are being produced every single second enormously and exponentially. IoT data exhibits several distinct characteristics that differentiate it from traditional data sources commonly known as the 3V or three dimensions called Volume, Velocity and Variety [32] and now the concept of big data has expanded to include three additional dimensions: Veracity, Validity, and Value. These dimensions, along with the original three (Volume, Velocity, and Variety), form the "6 Vs" framework for understanding big data [33], as illustrated in Figure 1.4. The 1st V stands for Volume: IoT generates large volumes of heterogeneous real time data due to the proliferation of connected devices and sensors. The sheer scale of data generated by IoT devices presents challenges in terms of storage, processing, and analysis.

The second V is Velocity; it pertains to the real-time or near real-time generation of IoT data, often at high speeds. . This characteristic requires systems and technologies that can handle the rapid influx of data and enable timely processing and decision-making. The third V, Variety, encompasses the diverse formats and structures of IoT data, ranging from structured, semi-structured, to unstructured data. This includes sensor readings, images, videos, audio, text, and more. Effectively managing and extracting insights from this variety of data types requires flexible data processing and analysis techniques. Veracity, regarded as a crucial 4th dimension, involves the trustworthiness of the received information[33]. IoT data quality may vary due to factors such as sensor inaccuracies, data transmission errors, or noise introduced during data collection. Ensuring data veracity involves data validation, cleansing, and quality control processes to enhance data reliability and accuracy. Value, as the fifth dimension among the V's, represents the ultimate goal of collecting IoT data. The primary objective is to extract actionable insights and generate value. IoT data has the potential to reveal patterns, trends, and correlations that can drive optimization, enhance efficiency, and enable informed decision-making. Variability, regarded as the final V among the Vs, emerges from the temporal and contextual fluctuations observed in IoT data. The data generated by IoT devices has the potential to undergo changes based on environmental conditions, device interactions, user behavior, or system states. By fully harnessing the complete potential of IoT data, organizations can unlock new opportunities and enhance decision-making capabilities across various domains.

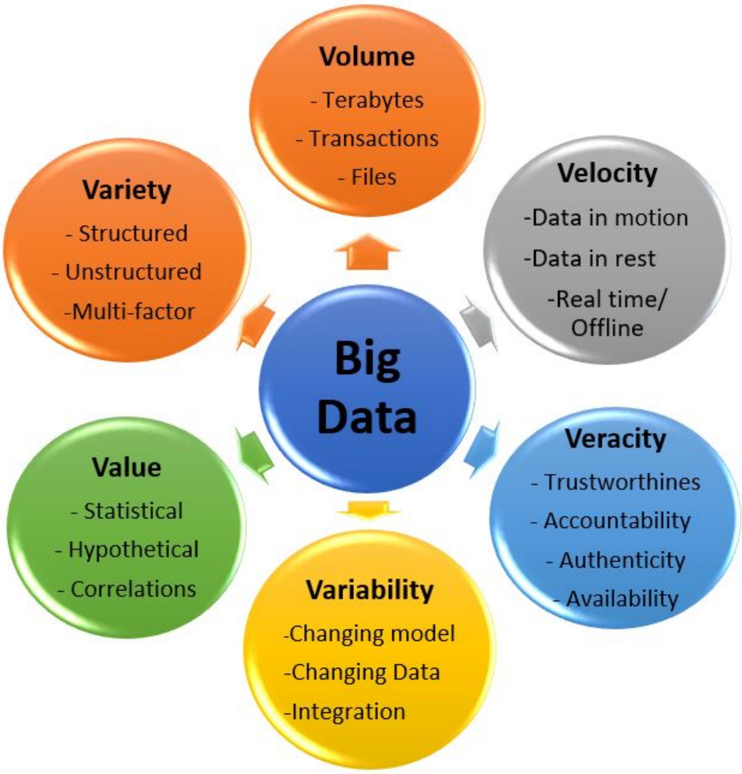


Figure 1.4 – *Big Data 6Vs*

1.4 Data Processing and analysis in IoT

IoT is a network of interconnected devices and objects embedded with sensors, software, and other technologies that enable them to collect and exchange data over the internet. The raw data generated from IoT environments are massive, diverse, continuous, and subject to errors, they might be uncertain, imprecise, or even erroneous, making effective data processing essential to extract valuable insights and enable intelligent actions. Data processing plays a crucial role in turning raw data into actionable knowledge, enabling informed decision-making and problem-solving in various fields and industries, data processing in IoT typically involves the collection, storage, analysis, and utilization of data generated by IoT devices and sensors.

1.4.1 Data processing cycle

Data processing in IoT adheres to the conventional Input>Process>Output cycle observed in various computer activities as shown in figure 1.5. This cycle involves collecting data from IoT devices (Input), performing necessary operations and analysis on the data (Process), and finally presenting valuable insights or triggering actions based on the processed data (Output). It should be noted that there is a difference between data and information. Data represent raw, unorganized facts that are generally of limited use until they undergo processing. Once data are subjected to processing, they transform into meaningful and valuable information useful for decision-making and understanding.

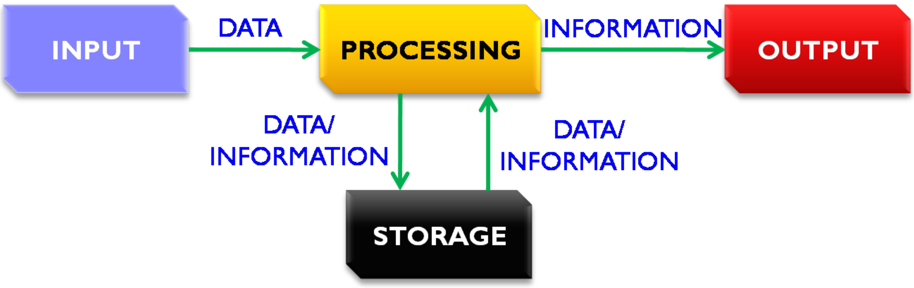


Figure 1.5 – *Data processing cycle*

1.4.1.1 Input

Input is the initial stage of the data processing cycle in IoT, and it is crucial for transforming the collected raw data into a format that can be understood and processed by computers or IoT systems. The data collected from various IoT devices can come in different forms, such as images, QR codes, text, numerical values, videos, and more. Before further processing, this diverse data must undergo conversion into a machine-readable format. This transformation ensures that the data can be efficiently analyzed, interpreted, and used to generate meaningful insights and actions as part of the IoT data processing workflow.

1.4.1.2 Processing

The processing stage is where the real value of IoT data is unlocked. Once these data are input into the system, they undergo various operations and transformations to convert them into meaningful information. This may involve filtering, summarizing, classification fusion and other data manipulation techniques, enabling the IoT system to extract valuable insights, identify patterns, detect anomalies, make predictions, and perform other data-driven operations.

1.4.1.3 Output

The processed data is presented to the users or applications in a human-readable format. This stage involves generating reports, charts, graphs, tables or any other form of output that conveys valuable information from the data, allowing end-users and decision-makers to understand and interpret the results of the data analysis easily, to make informed decisions, take necessary actions, and gain valuable insights into the performance and behavior of the IoT system. Additionally, the output data can be stored for further analysis and future reference which is essential for understanding trends, patterns, and changes over time.

As previously mentioned, many techniques are available to process data. However, this thesis will specifically focus on data fusion analysis, highlighting its crucial role in enhancing the quality and usefulness of processed data.

1.4.2 Basic architecture

The foundational architecture for processing IoT sensor data involves distinct layers, Iot data layer, data processing layer, data fusion and data analysis layer. At the core of this structure lies the IoT sensor data layer, comprising various devices capable of measuring physical parameters and capturing real-time changes in the environment. Among the commonly employed IoT sensors are those designed for temperature, pressure, humidity, level, accelerometer, gas, gyroscopes, motion, image, optical data, Radiofrequency Identifier (RFID), and Infra-Red (IR). These sensors are intricately linked with essential components such as the microprocessing unit, storage unit, control unit, power system, and wireless communication interfaces.

IoT sensor devices, however, operate within constraints related to size, computing power, memory, networking capabilities, and storage space. To facilitate communication, wireless protocols such as Wi-Fi, Zig Bee, Bluetooth, Near Frequency Communication (NFC), and LTE/4G are commonly employed in the communication interfaces of IoT sensor devices. This comprehensive architecture depicts the intricate integration of various components essential for processing, fusing, and analyzing data from IoT sensors. The data processing layer is directed towards various functions, including but not limited to data denoising, imputing missing data,detecting data outliers and aggregating data. Data fusion layer serves as a bridge between data processing and data analysis layers, contributing to both the preparation of integrated datasets and the extraction of valuable information from them.

In the context of data processing, data fusion involves the technical aspects of combining and refining raw data from various sources. This can include tasks such as cleaning, filtering, and integrating data to create a unified dataset. On the data analysis side, data fusion goes beyond the technical aspects and involves the interpretation and extraction of meaningful insights from the integrated dataset. It often employs statistical methods, machine learning techniques, or other analytical approaches to derive knowledge from the combined information.

The data analytic layer is concerned with extracting meaningful insights and knowledge from processed data using different approaches including, deep learning, machine learning and artificial intelligence, this layer also incorporates various intelligent functionalities, including cloud computing, fog computing and edge computing, to reduce computation and storage costs, enhance the network transmission reliability and improve IoT network security and privacy. These functionalities aim to meet the diverse needs of IoT-based applications. Figure 1.6 summarizes the basic architecture for IoT data processing, fusion and analysis layers.

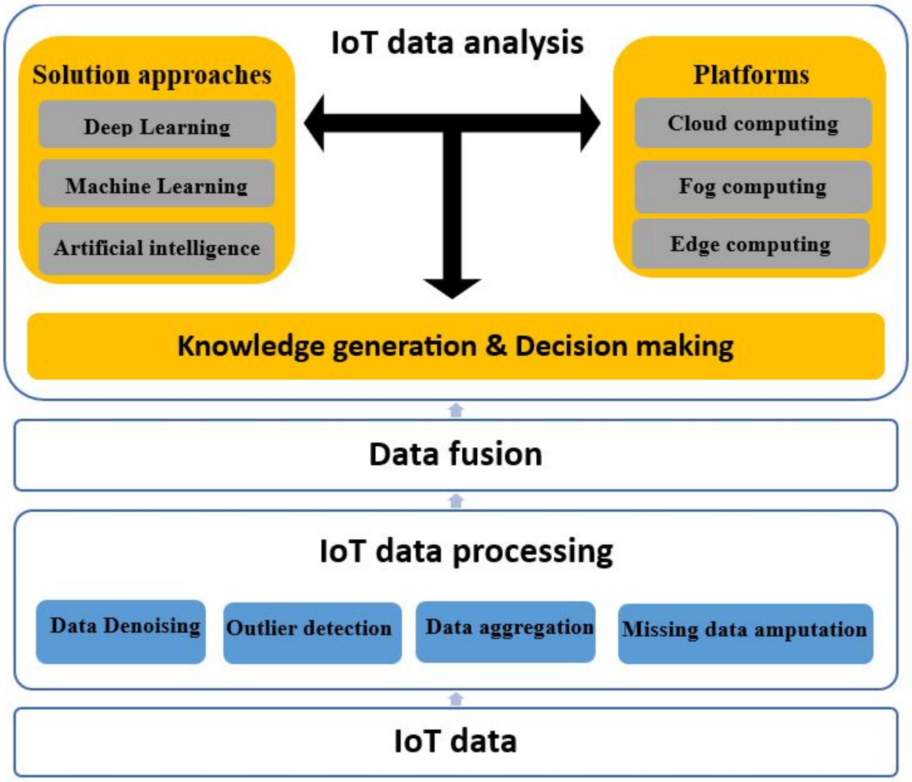


Figure 1.6 – *Basic architecture of data processing [34]*

1.5 Data fusion

1.5.1 Overview

Data fusion is the process that involves merging data from various sources to provide a robust and complete description of an environment or process of interest[35]. This method mimics the way the human brain integrate inputs from different senses, such as vision, hearing, touch, and more to form a unified and coherent perception of the environment.

Data fusion is of special significance when dealing with substantial volumes of data that need to be combined, fused and refined to obtain information of appropriate quality and integrity. The resulting synthesized data provides high-quality and trustworthy insights that serve as a foundation for making informed decisions.

Multisensor data fusion enhances accuracy and reliability by leveraging the strengths of each sensor while mitigating their weaknesses. Unlike relying on a single sensor, which may be subject to various sources of error or noise, combining data from multiple sensors results in more precise and dependable observations.

Data fusion is an old data processing technique that emerged in the 1960s as a mathematical method for data manipulation, notably with the introduction of the Kalman filter [1]. Its applications expanded in the 1970s when the US Department of Defense started utilizing this technology for defense and monitoring purposes, particularly in the military domain for tasks like tracking and airspace surveillance[36]. Over time, data fusion has found widespread deployment and implementation in various domains. It has been employed in the field of image processing[37]. Robotics has also benefited from data fusion, with applications in areas such as navigation and sensor integration [38, 39]. Medical applications have utilized data fusion for tasks like patient monitoring and diagnosis [40, 41]. Aerospace industry has employed data fusion techniques for tasks like target tracking and situational awareness [42, 43].

Data fusion is not limited to specific domains but is also used in distributed and heterogeneous environments such as smart IoT environments [44, 45]. In IoT, data fusion is employed in various applications, including smart home systems [46, 47], smart healthcare for remote patient monitoring and healthcare management [48], smart transportation for traffic monitoring and optimization [49], and smart agriculture for precision farming and crop monitoring [50].

Overall, data fusion has evolved from its inception as a mathematical technique and has found extensive utilization in diverse domains and environments for enhancing decision-making, improving situational awareness, and extracting valuable insights from multiple data sources.

Multisensor Data fusion process typically includes four primary steps: modelling, estimation, combination, and decision[51]; (i) Modeling: This step involves representing and describing the information collected using a fusion formalism. It aims to create a structured representation of the data to be fused.(ii) Estimation: In this step, the numerical distributions necessary for estimating the information to be fused are determined. It entails analyzing the available data and deriving the appropriate statistical models or parameters. (iii) Combination: The combination step is crucial as it focuses on selecting the appropriate operator or method to merge the data. (iv) Decision:The final step involves selecting the most appropriate decision criterion to enhance and merge the information effectively. This criterion relies on predefined rules or algorithms that take into account the objectives and requirements of the fusion application, ensuring that the fused information meets the desired outcomes.

Multisensor data fusion technique is more effective than relying on a single sensor because it provides more accurate and reliable data. This is because different sensors may have different strengths and weaknesses, and may be subject to different sources of error or noise. By combining data from multiple sensors, the system can leverage the strengths of each sensor and mitigate the weaknesses, resulting in a more accurate overall observation. The utilization of data fusion proves to be more advantageous compared to relying solely on a single sensor, as it facilitates the acquisition of more precise and dependable data. This advantage arises from the fact that different sensors possess distinct strengths and limitations, and are susceptible to varying sources of errors or noise. Through the combination of data obtained from multiple sensors, the system can effectively harness the individual strengths of each sensor while mitigating their weaknesses. As a result, a more accurate and comprehensive observation is achieved, enhancing the overall quality and reliability of the data and enabling more informed and effective decision-making.

1.5.2 Definitions

In the literature, multiple definitions of data fusion can be found, encompassing various aspects and emphasizing different facets of the data fusion process. Castenado [52]defined data fusion as "a combination of multiple sources to obtain improved information, which may be less expensive, of higher quality, or more relevant." Bostrom et al. (2007) provided another definition, stating that "Information fusion is the study of efficient methods for automatically or semi-automatically transforming information from different sources and different points in time into a representation that provides effective support for human or automated decision making."

Jitendra R. Raol [53]proposed a definition of multi-sensor data fusion as "the process of combining or integrating measured or pre-processed data or information originating from different active or passive sensors or sources to produce a more specific, comprehensive, and unified dataset or world model about an observed entity or event of interest." Mitchell [54] offered the widely recognized definition of data fusion as "the theory, techniques, and tools used for combining sensor data or data derived from sensory data into a common representational format."

Llinas and Hall [1]described data fusion as the process by which "data fusion techniques combine data from multiple sensors and related information from associated databases to achieve improved accuracy and more specific inferences compared to using a single sensor alone." Finally, the Joint Directors of Laboratories [55]provided an extensive and widely accepted definition, stating that data fusion is "a multi-level process that deals with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position, identify estimates, and provide complete and timely assessments of situations, threats, and their significance."

All these definitions fall within the same context. Data fusion refers to the combination of data gathered from heterogeneous sources in the most effective manner to extract the most useful information, providing a consistent and accurate representation that aids in obtaining a unified picture of a situation of interest.

1.5.3 Data Fusion in IoT applications

Data fusion in IoT has permeated virtually every aspect of our lives and can be observed across a multitude of applications and industries. As the IoT ecosystem expands and diversifies, data fusion gains even greater significance owing to its remarkable capacity to deliver invaluable insights and enhance a wide array of processes. This data integration and synthesis enable disparate IoT devices to collaborate effectively, generating a more comprehensive understanding of the interconnected world. By combining data from multiple sources, data fusion empowers businesses, organizations, and individuals to make data-driven decisions, streamline operations, and unlock unprecedented efficiencies. As a pivotal force driving the evolution of IoT technology, data fusion continues to shape the way we interact with our surroundings, ultimately contributing to a more intelligent, interconnected, and optimized world. Here are several instances showcasing the utilization of data fusion in IoT across various domains:

1.5.3.1 Military

The military domain has been at the forefront of adopting and driving the development of various technologies, including Wireless Sensor Networks (WSNs) and data fusion techniques. WSNs in the military context refer to a network of sensors that communicate wirelessly to gather data and provide critical information for surveillance, defense, and intelligence purposes.

Data fusion, as applied in the military, involves integrating and analyzing data from diverse sensors, platforms, and sources to enhance situational awareness, decision-making, and operational effectiveness. By combining data from sources such as unmanned aerial vehicles (UAVs), ground sensors, satellite imagery, and command-and-control systems, military forces can obtain a comprehensive understanding of the battlefield environment.

1.5.3.2 Healthcare

Data fusion in IoT plays a vital role in the modern healthcare landscape, where a vast amount of data is generated from various sources, including wearable devices, electronic health records (EHRs), medical imaging, mobile applications, and more. enabling remote patient monitoring, real-time health data collection, and analysis.

1.5.3.3 Transportation

Data fusion in IoT for transportation involves the integration and analysis of data from various IoT devices and sensors within the transportation ecosystem. It aims to optimize transportation operations, improve safety, enhance traffic management, and provide better services to commuters. IoT data fusion in transportation combines data from vehicles, infrastructure, traffic signals, weather, and other relevant sources to gain valuable insights and make informed decisions.

1.5.3.4 Agriculture

Data fusion in IoT for agriculture involves leveraging interconnected devices and sensors to collect, integrate, and analyze data across various aspects of agricultural operations. By combining data from sources such as soil moisture sensors, weather stations, crop health monitors, and equipment trackers, farmers gain comprehensive insights into their fields' conditions and crop growth patterns. This integrated approach enables farmers to make data-driven decisions regarding irrigation scheduling, pest management, fertilizer application, and overall farm management.

1.5.3.5 Aeronautics

Multisensor data fusion in IoT for aeronautics enables enhanced situational awareness, improved navigation and overall accurate and comprehensive insights into the operational status, performance, and aircraft safety. The applications of IoT multisensor data fusion for aeronautics are diverse and include:

- **Aircraft Health Monitoring**: By combining data from various onboard sensors measuring parameters such as engine performance, airspeed, altitude, temperature, and structural integrity, engineers can monitor the health of aircraft systems in real-time. This allows for predictive maintenance, early detection of faults, and enhanced safety.

- **Flight Optimization**: Multisensor data fusion enables the optimization of flight paths and fuel consumption by integrating information from weather sensors, air traffic control systems, and aircraft performance metrics. This helps airlines reduce costs, minimize environmental impact, and improve overall operational efficiency.
- **Collision Avoidance Systems**: Integrating data from radar, lidar, and other proximity sensors enables the development of advanced collision avoidance systems that enhance airspace safety. These systems can detect and mitigate potential conflicts between aircraft, as well as obstacles such as terrain and buildings.
- **Environmental Monitoring**: IoT sensors deployed on aircraft can collect data on atmospheric conditions, pollution levels, and other environmental factors during flight. Multisensor data fusion techniques allow for the aggregation and analysis of this data to support research on climate change, air quality, and other environmental concerns.
- **Remote Sensing and Surveillance**: Multisensor data fusion enables the integration of data from drones, satellites, and other remote sensing platforms with data from onboard aircraft sensors. This facilitates applications such as aerial imaging, agricultural monitoring, and disaster response.

1.5.4 Multisensor Data fusion architecture

The architecture of multisensor data fusion refers to the whole process of multisensor data fusion, including the constituent components of the fusion system, the primary functions performed by each component, the interconnections between these components, the relationships between subsystems and the overall system, the fusion location, and other related aspects[56]. In the mid-1980s, the Joint Directors of Laboratories (JDL) formed the Data Fusion Subpanel, which later became known as the Data Fusion Group (DFIG). It is a generalized framework commonly used to describe the process of data fusion. It provides a conceptual representation of the various stages of data fusion [55].The JDL model originally includes four processing levels, which are now adjusted to seven [57], each level corresponds to a specific aspect of the data fusion process as depicted in Figure 1.7 and explained as follows: :

Level 0: Source Preprocessing (or Data Assessment) This level involves the initial processing of raw sensor data, such as calibration, filtering, and artifact removal.

Level 1: Object Assessment Based on data and their relationships and state estimation, this level provides evaluation and prediction of entities and objects.

Level 2: Situation Assessment This level offers an assessment and forecast of relationships correlations, and patterns among objects.

Level 3: Impact assessment or Threat Refinement This level involves assessing the potential consequences and impact of identified threats. It considers factors such as vulnerability, criticality, and potential damage caused by the threats.

Level 4: Process Optimization This level involves resource management and adaptive information collection and processing. It forecasts the impact of planned or estimated actions by participants.

Level 5: User Refinement (or cognitive refinement) Managing knowledge, this level enables adaptive access to control and information display. It supports decision-making through human-machine interfaces.

Level 6: Mission Management (Mission Refinement) Aspects of managing systems, this level allows for the spatial-temporal management of resources, preparation, and target setting. It aids decision-making while considering societal, financial, and political constraints.

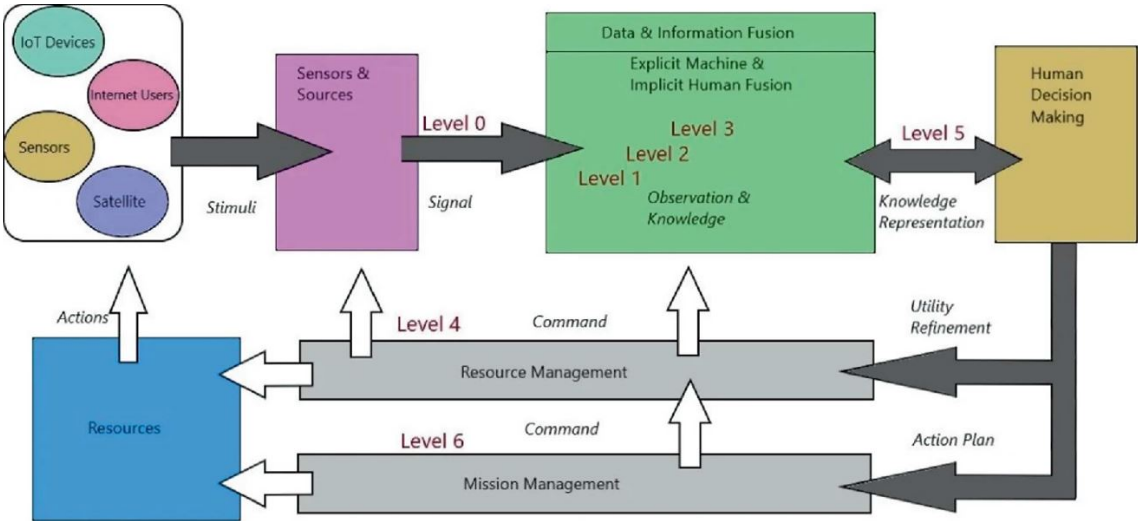


Figure 1.7 – *Joint Director Laboratories JDL model [57]*

1.5.5 Multisensor data fusion classification

Data fusion solutions can be systematically classified, drawing on diverse criteria including the nature of the data being fused, the intended purpose of the fusion process, the specific techniques...etc. This classification primarily focuses on the connection between data types and processing levels, and it can be described as follows:

1.5.5.1 Abstraction levels

Data fusion technique can be performed at various levels, depending on the representation of data to be merged and according to the stage at which the combining operation takes place. This classification consists of four levels: low, intermediate, high and multilevel fusion as follows[58]:(see Figure 1.8 :

1. **Low level fusion or raw data fusion** The raw data and measurements obtained from various sources are combined and analyzed directly, without undergoing any preprocessing, to generate a new raw data representation that is more informative and comprehensive than what could be obtained from individual sources. The fusion process focuses on integrating and aligning the raw data streams, removing redundancies, and enhancing the quality of the data.

2. **Intermediate level or feature level fusion**

   It takes place after the raw data has been pre-processed, and features are extracted from each data source. The extracted features capture relevant information and characteristics of the data. The extracted features from different sources are combined to create a new feature set that is more useful and precise.

3. **High level or decision level fusion**

   It occurs at a higher level of abstraction, and the result of the fusion is the basis for command and control decision-making. In this level of fusion process, the inputs from various sensors are first analyzed and processed separately, the results obtained from each individual fusion process are then combined to generate a final result or a more accurate and reliable decision.

4. **Multilevel fusion:**

   It involves combining raw data, features, and/or decisions to generate an output at a specific level.

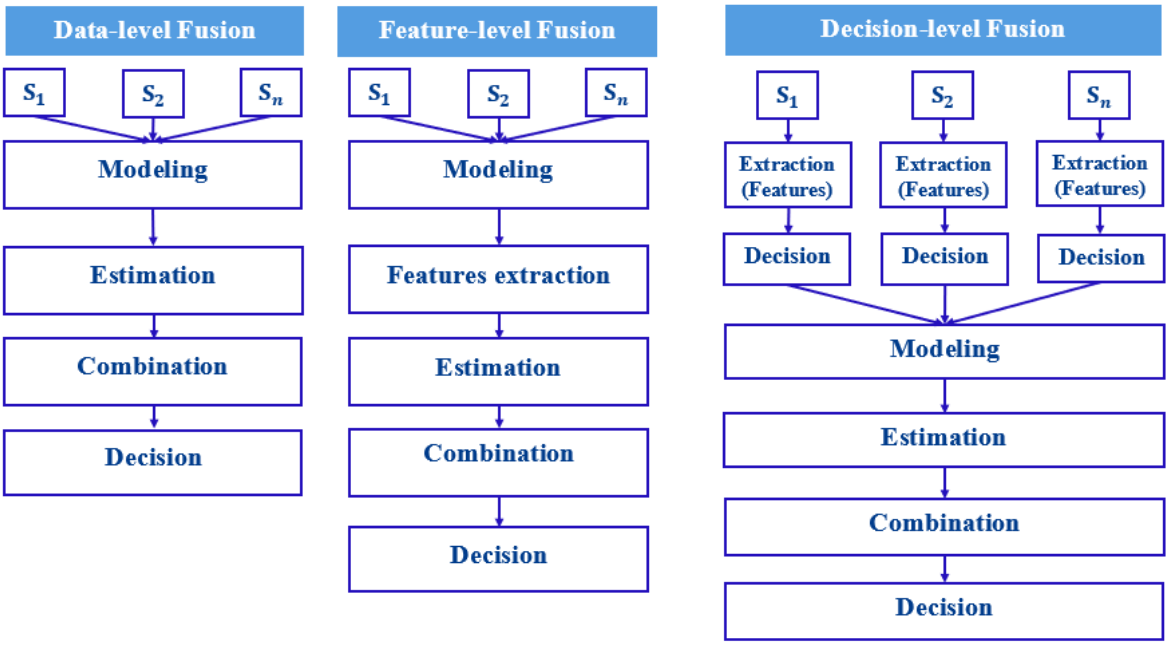


Figure 1.8 – *Data fusion levels*

1.5.5.2 Sensor configuration

Based on the relationship among data sources, data fusion can be classified into three levels [59]:

1. **Complementary fusion**:Two or more data inputs from different parts of the same target are combined to achieve a more comprehensive and holistic understanding of the information.

2. **Redundant fusion**: Data obtained from multiple sources, which essentially represent the same information, are merged to enhance the quality and reliability of the data.

3. **Cooperative fusion**: Data from independent sources are integrated to generate new data or obtain more intricate and advanced information.

1.5.5.3 Input/Output relations (Dasarathy's Classification)

It considers the relationships between different data types in the input and output of fusion systems. These relationships are constrained by the level of data, where the output level must be higher than the input level. This classification encompasses six abstraction levels [60]:

1. **Data In-Data Out (DAI-DAO)** This is the basic level of data fusion, Raw data are used as input, i.e; the fusion occurs immediately after data collection from sensors and it generates more reliable and/or accurate data as output.

2. **Data In-Feature Out (DAI-FEO)** This level involves extracting features or characteristics that describe entities in the environment from raw data as the output.

3. **Feature In-Feature Out (FEI-FEO)** In this level, both the input and output of the fusion process are features,where existing features are improved or new ones are extracted.

4. **Feature In-Decision Out (FEI-DEO)** This level takes a set of features as input and provides a set of symbolic representations or decisions as the output.

5. **Decision In-Decision Out (DEI-DEO)** This level combines input decisions to obtain more reliable or new decisions as the output.

6. **Temporal (data/feature/decision) fusion** Integration of different data over various time periods, which can be applied at any level, often used for tracking purposes.

1.5.5.4 Time vector and space vector

Data fusion can be categorized into three types based on the time vector and space vector:

1. **Time fusion** involves the fusion processing of time-domain data from a specific sensor in the system.

2. **Spatial fusion** refers to the fusion processing of measurement values from related targets at the same sampling time for each sensor in the system.

3. **Spatiotemporal fusion** involves the fusion processing of measurement values from relevant targets of the sensors in the system over a period of time

1.5.5.5 Fusion data attributes

Based on the attributes of fusion data, multisensor data fusion can be categorized into homogeneous data fusion and heterogeneous data fusion.

1. **Homogeneous data fusion** involves the consistent representation of fusion processes using homogeneous data collected by multiple identical sensors. It is also known as multisensor homogeneous data fusion.

2. **Heterogeneous data fusion**, on the other hand, refers to the process of achieving a consistent representation of fusion using heterogeneous data collected by multiple different sensors. This category is also known as multisensor heterogeneous data fusion.

1.5.5.6 Data fusion architecture

Depending on the type of architecture, data fusion can be classified into three categories[61]:

1. **Centralized architecture** involves a merge node residing in the central processor, which receives information from all input sources in the form of metrics.

2. **Decentralized architecture** refers to each node fusing its local information with the data received from its peers.

3. **Distributed architecture** entails processing measurements from each source node independently before sending the information to the fusion node.

1.5.5.7 Data fusion techniques

Data fusion solutions can be broadly classified into three fundamental categories, considering the employed method [1] (see Figure 1.9)

- **Probability-based methods:** Probability-based methods utilize the power of probabilistic models and statistical techniques to integrate diverse data effectively. They embrace (i) Bayesian analysis, which allows for the updating of probabilities based on new evidence, (ii) Statistical tools, such as regression analysis for exploring relationships between variables, hypothesis testing for validating assumptions, and estimation theory for accurately estimating unknown parameters. (iii) Recursive operators, such as Kalman filter, offer a dynamic approach to continually refine and update fused data, making probability-based methods vital for scenarios where uncertainty is inherent.

- **Artificial Intelligence (AI) based techniques:** AI-based data fusion techniques leverage advanced algorithms and approaches to extract meaningful patterns from disparate data. Classical machine learning methods, including decision trees, support vector machines, and random forests, excel in uncovering complex relationships and making predictions. Fuzzy logic, which accommodates imprecise information, and artificial neural networks (ANNs), capable of learning complex non-linear relationships, contribute to the adaptability of AI-based techniques. Genetic algorithms, inspired by natural selection, enhance optimization and pattern recognition. These techniques are particularly adept at handling large data and identifying intricate patterns that conventional techniques may fail to detect.

- **Theory of Evidence-based Data Fusion methods:** Rooted in the theory of evidence, these data fusion methods operate on the principles of the Dempster-Shafer theory also known as the theory of belief functions. This theoretical foundation provides a systematic approach to reasoning with uncertain, imprecise, and incomplete data. The theory allows for the combination of evidence from different sources while explicitly modeling uncertainty and conflict. By assigning belief values to hypotheses, evidence-based data fusion methods enable a nuanced representation of knowledge, making them valuable in situations where information may be contradictory or incomplete. This category offers a principled framework for handling uncertainty and enhancing the robustness of fused data.

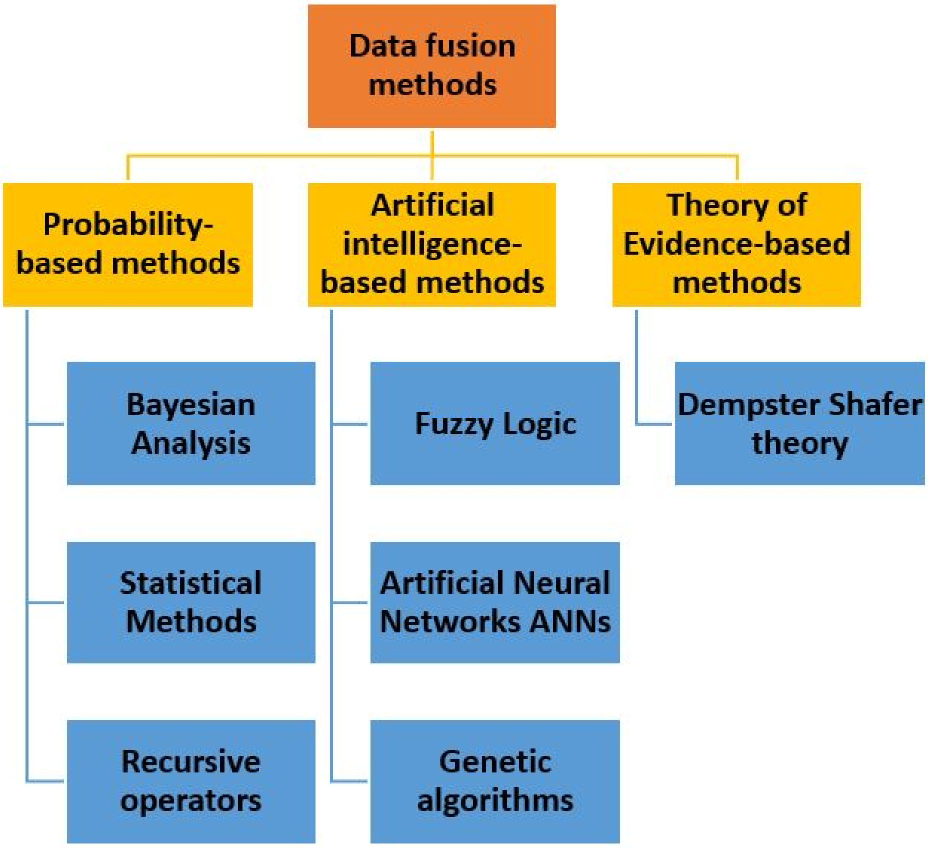

Data fusion methods
Probability-based methods
Artificial intelligence-based methods
Theory of Evidence-based methods
Bayesian Analysis
Statistical Methods
Recursive operators
Fuzzy Logic
Artificial Neural Networks ANNs
Genetic algorithms
Dempster Shafer theory

Figure 1.9 – *Data fusion techniques*

2

Mathematical methods for Data Fusion in IoT

Contents

2.1 Introduction

The process of data fusion requires the utilization of formalisms capable of effectively combining acquired data from diverse sources. However, fusion results can be erroneous due to various forms of imperfections inherent in the data, primarily stemming from imprecision, uncertainty, and conflict. These imperfections pose significant challenges to the fusion process, as they introduce ambiguity and inconsistencies that can affect the reliability and accuracy of the fusion results. Therefore, it is important to study the techniques that enable understanding and solving issues related to data fusion in the Internet of Things, ensuring the effectiveness and reliability of the data fusion process.

In this chapter, we examine the most relevant and well-established multisensor decision-level fusion techniques in the context of the Internet of Things (IoT), specifically the Bayesian method, fuzzy logic, and theory of belief functions that consider the different data imperfections. Simultaneously, we conduct a comparative study to discern the intrinsic characteristics of data fusion theories. This involves elucidating the primary rules employed in each method's various data fusion steps and highlighting the advantages and limitations associated with each approach. A particular focus is placed on addressing challenges related to data imperfections such as uncertainty, imprecision and conflict, as well as considering constraints imposed by IoT environments, including real- time requirements. Subsequently, we select the most suitable fusion formalism for IoT environments based on the acquired information.

2.2 Data imperfections

Data obtained from various sources often exhibits imperfections in different forms. These imperfections can be attributed to observed phenomena, sensor limitations, noise or lack of reliability, etc. The spectrum of data imperfections is diverse, with primary manifestations including imprecision, uncertainty, and conflicts.

Uncertainty

Uncertainty refers to the lack of certainty or confidence in the accuracy, reliability, or interpretation of data, stemming from inadequate knowledge or understanding of its true nature or value. It indicates a qualitative deficiency within the information itself, reflecting either partial or complete ignorance regarding the subject matter. Uncertainty may arise from various sources, including measurement errors, incomplete information, or inherent randomness in the data-generating process. Uncertainty can manifest in various forms and contexts. Two common types can be distinguished; aleatory uncertainty and epistemic uncertainty. The former is defined as "uncertainty due to inherent randomness" and the latter as "uncertainty due to lack of knowledge.

- **Epistemic uncertainty**: refers to uncertainty arising from a lack of knowledge or information about a system or phenomenon. This type of uncertainty can stem from limitations in data, incomplete understanding of underlying processes, or ambiguity in model assumptions. It is subjective and reducible through the acquisition of additional knowledge.
- **Random or aleatory uncertainty**: this type arises from inherent variability in a system characterized by its random parameters. It is irreducible and stochastic, and generally related to genuinely unpredictable events, such as natural disasters or quantum phenomena.

Imprecision

Imprecision refers to the lack of exactness or specificity in data. It occurs when data values or measurements are not precise or granular enough to accurately represent the underlying information. Imprecision can arise due to measurement errors, limitations of data collection methods, or inherent variability in the data itself. Dealing with imprecision involves understanding the level of uncertainty associated with the data and considering the potential impact on analysis and decision-making processes.

Conflicts

Conflicts emerge when inconsistencies, contradictions, or discrepancies exist between different data s or between data and existing knowledge or beliefs. Conflict can occur due to measurement errors, conflicting data sources, or different interpretations of the same information. Resolving conflicts involves identifying the sources of disagreement, reconciling conflicting information, and making informed decisions about the most reliable or accurate representation of the underlying reality.

2.3 Bayesian Approach

2.3.1 Introduction

The Bayesian approach stands as the oldest and the most mathematically well-developed method, offering a probabilistic framework for interpreting the concepts of chance and uncertainty. It forms an integral part of numerous data fusion techniques [62] due to its robust mathematical foundations. This approach uses probability distributions to express uncertain quantities and probability theory for combining data coming from heterogeneous sources.

In the Bayesian methodology, the process involves defining priors, specifying them, and engaging in posterior computations. The flexibility and adaptability of Bayesian inference make it a powerful tool for integrating information from diverse sources and making reliable inferences in the presence of uncertainty. These attributes have contributed to its widespread popularity across various data fusion applications [63, 64, 65].

2.3.2 Fusion process

2.3.2.1 Modeling

The modeling step is based on conditional probabilities as follows:

$$p(H_i/m_j) \tag{2.1}$$

Where H_i and m_j stand for a hypothesis and a measure respectively.
We assume $H_1 \ldots H_N$, a set of mutually exclusive assumptions that satisfy the following conditions:

$$\begin{cases} \forall\ i,j \ \ H_i \cap H_j &= \emptyset,\ i \neq j; \\ \cup_{i=1}^{N} H_i &= E \end{cases} \tag{2.2}$$

Where E represents the space of hypotheses.

2.3.2.2 Combination

In a Bayesian framework, Bayes' rule represents the law of combining several probability distributions, it is used to estimate the posterior probability of a hypothesis from the prior probabilities, i.e. the probability of the occurrence of a future event is estimated by observing the occurrence of similar events in the past.

Let m_1 and m_2 be two characteristic primitives from different sources representing the same hypothesis, Bayes' rule provides the possibility to evaluate the posterior probability of hypothesis H_i, knowing the measures m_1, m_2:

$$p(H_i/m_1, m_2) = \frac{p(H_i).p(m_1, m_2/H_i)}{\sum_{j=1}^{N} p(H_j).p(m_1, m_2/H_j)} \tag{2.3}$$

where:

- $p(H_i/m_1, m_2)$ is called the posterior probability.
- $p(m_1, m_2/H_i)$ is called the likelihood function and is based on the given sensor measurement model.
- $p(H_i)$ is called the prior distribution and incorporates the given transition model of the system.

- The denominator is a normalizing term to ensure that the probability density function integrates to one, it is known as "the evidence" and it is constant for all events.

2.3.2.3 Decision

For the decision phase, many criteria are possible, the most frequently used are the following:

- Maximum A Posteriori (MAP): it requires choosing the hypothesis with the greatest posterior probability.

$$P(H_k/m_1, m_2) = \max_{i \in \{1,\ldots,n\}} P(H_i/m_1, m_2) \tag{2.4}$$

- Maximum of Likelihood (ML): It is about choosing the hypothesis with the biggest likelihood probability. We choose H_k, if H_k is the solution of the following equation:

$$\frac{\partial L(m_1, m_2, H_i)}{\partial H_i} = 0 \tag{2.5}$$

With,

$$\frac{\partial^2 L(m_1, m_2, H_i)}{\partial H_i^2} < 0 \tag{2.6}$$

The likelihood function is given by:

$$L(m_1, m_2, H_i) = P(m_1, m_2, H_i) \tag{2.7}$$

- Maximum ENtropy (MEN): The decision is made by maximizing entropy as follows:

$$h(H_k/m_1, m_2) = \max_{i \in \{1,\ldots,n\}} h(H_i/m_1, m_2) \tag{2.8}$$

With,

$$h(H_i/m_1, m_2) = -P(H_i/m_1, m_2).log\ P(H_i/m_1, m_2) \tag{2.9}$$

- Maximum Expectation (MEX): This time, the expected value is maximized

$$E(H_k/m_1, m_2) = \max_{i \in \{1,\ldots,n\}} E(H_i/m_1, m_2) \tag{2.10}$$

2.4 Fuzzy logic

2.4.1 Introduction

Fuzzy logic, also known as fuzzy mathematics, provides a mathematical framework for handling imprecise and vague information. It was first introduced by L. Zadeh in 1965 [66] as a simulation of human recognition. Its main idea is to transmit human reasoning richness to a computer. This theory employs mathematical tools to characterize fuzzy concepts and expand the conventional notion of sets to include fuzzy sets.

By introducing the concept of a "membership function," which captures the degree of ambiguity associated with an element's belonging, this non-probabilistic technique allows for modeling partial membership in a class with vague boundaries. Thus, a given element can belong to a class with a degree of membership ranging from 0 to 1, enabling a flexible representation of uncertain or imprecise information. This modeling also enables the incorporation of symbolic information and knowledge expressed in natural language.

This approach overcomes the limitations faced by computers in handling fuzzy concepts, thereby providing a framework to model and analyze real-world complexities with greater precision, it has been widely deployed in various data fusion applications [67, 68, 69].

2.4.2 Theoretical foundations

2.4.2.1 Membership function

A fuzzy set is characterized by its membership function μ, which captures the gradual or partial membership of elements within the set. Unlike crisp sets with precisely defined boundaries, fuzzy sets exhibit ambiguous and vague boundaries. In this context, a universal set X, also referred to as the universe of discourse, consists of individual elements denoted as x. Each fuzzy subset in X is associated with a membership function $\mu_A(x)$, which assigns a real number between 0 and 1 to every element x in X. The value of $\mu_A(x)$ represents the graded membership of x in the fuzzy set A. Thus, the membership function is defined as:

$$\mu_A(x) : \mathrm{X} \to \Big[0, 1\Big] \tag{2.11}$$

And the fuzzy set can be expressed by a set of ordered pairs as follows:

$$A = \Big\{\Big(x,\ \mu_A(x)\Big),\ x \in X,\ \ 0 \leqslant \mu_A(x) \leqslant 1\Big\} \tag{2.12}$$

The membership function serves as the basis for fuzzy set theory, with determination relying on both experiential knowledge and statistical analysis. Currently, three primary types of membership functions are prevalent: Gaussian, triangular, and trapezoidal.

- **Gaussian membership function:**

$$\mu(x) = \exp \frac{-(x-c)^2}{2\sigma^2} \tag{2.13}$$

  Where $'c'$ denotes the mean value while $'\sigma'$ represents the standard deviation of the Gaussian membership function. This type exhibits smooth and stable transition characteristics.

- **Triangle membership function:**

$$\mu(x) = \begin{Bmatrix} 0, & x \leqslant f \\ \dfrac{x-f}{m-f}, & f < x \leqslant m \\ \dfrac{g-x}{g-m}, & m < x \leqslant g \\ 0, & g < x \end{Bmatrix} \tag{2.14}$$

  Where $'f'$ and $'g'$ stand for the abscissa values of the left and right vertices at the base of the triangle, while 'm' represents the abscissa value of the apex at the top of the triangle. The structure of this type is simple and easily calculable.

- **Trapezoidal membership function:**

$$\mu(x) = \begin{Bmatrix} 0, & x \leqslant f \\ \dfrac{x-a}{b-a}, & a < x \leqslant b \\ 1, & b < x \leqslant c \\ \dfrac{d-x}{d-c}, & c < x \leqslant d \\ 0, & x > d \end{Bmatrix} \tag{2.15}$$

  Where $'b'$ and $'c'$ denote the abscissas of the two vertices forming the upper base of the trapezoid, while $'a'$ and $'d'$ represent the abscissas of the two vertices forming the base of the trapezoid.

2.4.3 Fusion process

2.4.3.1 Fuzzification

Fuzzification is the process of converting crisp (non-fuzzy) inputs into fuzzy variables by assigning them membership grades or values within fuzzy sets. Membership functions can take various forms, such as triangular, trapezoidal, Gaussian, or sigmoidal shapes, depending on the nature of the input variable and the desired fuzzification.

2.4.3.2 Inference (Fuzzy rules)

Inference using fuzzy rules is a methodology that allows the representation of human knowledge, reasoning, or expertise using IF-THEN rules and fuzzy statements. There are two commonly used models for fuzzy inference:(i) the Mamdani model (also known as the fuzzy linguistic model) and (ii) the Takagi-Sugeno model.

- **The Mamdani model [70]** uses fuzzy propositions with linguistic variables that take on linguistic values instead of precise numerical values. For instance, instead of using a numerical value to represent temperature, linguistic variables such as "hot," "cold," or "warm" can be used. The IF-THEN rules in the Mamdani model involve fuzzy propositions and linguistic variables. These rules define the relationship between the inputs and outputs of a system and can be expressed in the form of "IF [antecedent], THEN [consequent]."

- **The Takagi-Sugeno model [71]** is based on numerical variables rather than linguistic variables. In this model, the fuzzy rules are expressed using numerical values and mathematical functions. The IF-THEN rules in the Takagi-Sugeno model are typically defined as "IF [conditions], THEN [consequent]," where the conditions are numerical and the consequent is a mathematical expression involving the input variables.

2.4.3.3 Composition

Fuzzy logic principles are used to handle imprecise or uncertain information. These principles involve combining IF-THEN rules from a fuzzy rule base to create a mapping from fuzzy input sets to fuzzy output sets. The rules are interpreted as fuzzy implications, and various operators are used to perform the necessary computations.

- **T-NORME operators** also known as triangular operators, exhibit a conjunctive behavior and combine information in a "logical AND" way. The most common T-NORM operator is the min-operator, which can be expressed as follows:

$$\mu_{A\cap B} = \min\Big(\mu_A(x), \mu_B(x)\Big) \;\; \forall\, x \in X \tag{2.16}$$

- **T-CONORME operators** or triangular conorm operators, exhibit a disjunctive behavior and combine information in a "logical OR" way. The most common T-CONORM operator is the max-operator, which can be expressed as follows:

$$\mu_{A\cup B} = \max\Big(\mu_A(x), \mu_B(x)\Big) \;\; \forall\, x \in X \tag{2.17}$$

- **Complementary operators:** represent a "logical NOT" operation and are used to negate the membership degrees of fuzzy sets. For example, the complementary operator for fuzzy set A can be defined as:

$$NOT\ A = \overline{A} = \left\{x,\ \mu_{\overline{A}}(x) | x \in X,\ \mu_{\overline{A}}(x) = 1 -\ \mu_A(x)\right\} \tag{2.18}$$

2.4.3.4 Defuzzification

Defuzzification is the process of converting fuzzy outputs obtained from fuzzy logic inference methods into crisp (non-fuzzy) outputs. It involves selecting a single value that best represents the fuzzy set. There are several forms of defuzzification including:

- **Center of gravity method (COG):** The COG method calculates the crisp value based on the center of gravity of the fuzzy set. The defuzzification value denoted as x^* using COG is defined as:

$$x^* = \frac{\int x\mu_A(x)dx}{\int \mu_A(x)dx} \tag{2.19}$$

- **Bisector of area method (BOA):** The BOA method determines the position under the curve where the areas on both sides are equal. It calculates the value x^* such that the area under the membership function from the minimum value α to x^* is equal to the area from x^* to the maximum value β. Mathematically, it can be expressed as:

$$\int_{\alpha}^{x^*} \mu_A(x)dx = \int_{x^*}^{\beta} \mu_A(x)dx \tag{2.20}$$

  With,

$$\alpha = \min\left\{x | x \in X\right\} \tag{2.21}$$

  And,

$$\beta = \max\left\{x | x \in X\right\} \tag{2.22}$$

- **Mean of Maximum Method (MOM):** The mean of maximum method calculates the defuzzified value by taking the average of all the input values that correspond to the maximum membership degree. It considers all the points at which the membership function reaches its maximum value and calculates their average as the defuzzified output.

$$x^* = \frac{\sum_{i=1}^{K} x_i}{k} \tag{2.23}$$

  Where k is the number of elements of the output fuzzy set that reach the maximum memberships.

2.5 Theory of belief functions (Dempster-Shafer theory)

2.5.1 Introduction

Theory of belief functions (Dempster-Shafer theory or evidence theory) is a formal framework for calculation, modeling and reasoning under uncertainty and imprecision, it allows for the manipulation of finite sets within a robust mathematical framework through the use of functions called belief functions that operate within the frame of discernment.

The theory was first initiated by A.P Dempster in 1967 [2] as a generalization of Bayesian inference, in his work which deals with lower and upper probability distributions, then developed by his student Shafer in 1976 [3] into a general framework of uncertain reasoning, by introducing the concept of "trust function" and adding the possibility of modeling the unknown knowledge. Unlike probability theory, D-S theory doesn't only allow for the allocation of a probability mass to mutually exclusive singletons, but also to sets or intervals and it doesn't require prior knowledge to combine the pieces of evidence.

To use Dempster-Shafer theory for multisensor data fusion, the first step is to define a set of hypotheses or states that the system being monitored can be in. Each source then provides evidence in support of or against these hypotheses. The evidence is represented using belief functions, which are functions that assign a degree of belief to each hypothesis. Once the belief functions have been defined for each source, they are combined using a rule called the Dempster's combination rule. This rule takes into account the degree of overlap between the evidence provided by each source and produces a new belief function that represents the combined evidence. The final step is to use the combined belief function to make a decision about the state or condition being monitored. This decision can be made using a variety of methods, such as selecting the hypothesis with the highest degree of belief or computing the expected value of a utility function.

Dempster Shafer theory has been extensively applied in various multisensor data fusion applications such as decision making [4, 5, 6], fault diagnosis [7, 8, 9], target recognition [10, 11, 12] . . . etc., owing to its flexibility and effectiveness in handling uncertainty problems and its ability in merging heterogeneous data obtained from multiple sources without prior knowledge.

2.5.2 Theoretical foundations

2.5.2.1 Frame of discernment

Denoted by Ω, it's a finite, nonempty set of mutually exclusive and exhaustive hypotheses, it is expressed as follows:

$$\Omega = \{A_1, A_2, \dots, A_n\} \tag{2.24}$$

The power set of Ω, denoted by 2^{Ω}, is the set of all possible subsets of Ω. For any $A \subseteq \Omega$, it is defined as:

$$2^{\Omega} = \{\emptyset, \{A_1\}, \ldots, \{A_n\}, \{A_1 \cup A_2\}, \ldots, \Omega\} \tag{2.25}$$

2.5.2.2 Basic Probability Assignment BPA or Mass function

It represents how strongly the evidence supports a hypothesis by assigning probability to the different subsets. In a frame of discernment, the mass function of a subset symbolized by m is defined as:

$$m_j : 2^{\Omega} \to [0, 1] \tag{2.26}$$

Satisfying the following conditions:

$$\begin{cases} m(\emptyset) = 0 \\ \sum_{A \subseteq \Omega} m(A) = 1 \end{cases} \tag{2.27}$$

The first condition stipulates that no belief mass should be allocated to the null set, essentially assuming that the actual value of Ω is a part of the universal set Ω, a principle referred to as the closed-world assumption.

The second condition states that the sum of the belief mass functions over all the subsets is unity, implying that the total belief is normalized to a measure of one. This normalization facilitates the allocation of belief to different propositions.

$\forall\ A \subseteq \Omega,\ if\ m(A) >\ 0$, A is called a focal element of evidence.

The mass functions corresponding to simple hypotheses express the certainty of a class compared to others, whereas the mass functions corresponding to compound hypotheses express the confusion arising from the lack of information to decide between one class or another.

2.5.2.3 Mass functions' determination

Estimating mass functions is indeed a delicate and crucial problem in Dempster-Shafer Theory. There are no universal mathematical methods to accomplish this step. Instead, the choice of method depends on the problem at hand and the available information. One frequently used method is subjective assessment, which relies on expert judgment or subjective assessment to assign evidence to different propositions based on available knowledge. However, other methods exist for estimating mass functions, each suited to different scenarios and data types:

- **Probabilistic models:** These models estimate belief mass based on conditional probabilities or probabilistic reasoning techniques such as the Gaussian distribution.
- **Distance models:** Distance-based methods, such as the k-nearest neighbor algorithm proposed by Denoeux [72] , estimate mass functions by considering the similarity or proximity between different pieces of evidence. These methods often rely on distance metrics to quantify the similarity between evidence sources or propositions, and they assign belief mass accordingly.
- **Estimation of mass functions on fuzzy subsets:** In this category, methods focus on estimating mass functions on fuzzy focal elements to model the fuzzy uncertainty inherent in the processed information. The methods for estimating these mass functions may vary depending on the specific application and the nature of the fuzzy evidence.

2.5.2.4 Uncertainty representation

Based on the definition of BPAs, Belief function (Bel) and Plausibility function (Pl) which represent-respectively-the lower and upper bounds of the uncertainty interval are defined as follows:

- **Belief function**: (Credibility) it represents the total belief in hypothesis A to be true. It is indicated by:

$$\begin{cases} Bel: \; 2^{\Omega} \; \rightarrow [0,1] \\ \text{Bel(A}_i) = \sum_{A_j \subseteq \; A_i} m \; (A_j) \end{cases} \tag{2.28}$$

  With,

$$\begin{cases} Bel(\emptyset) = 0 \\ Bel(\Omega) = 1 \end{cases} \tag{2.29}$$

- **Plausibility function**: it refers to the possible belief in the hypothesis A. It is defined as:

$$\begin{cases} Pl: \; 2^{\Omega} \; \rightarrow [0,1] \\ Pl(A_i) \; = \; \sum_{A_j \cap A_i \neq \emptyset} m(A_j) \end{cases} \tag{2.30}$$

  With,

$$\begin{cases} Pl(\emptyset) = 0 \\ Pl(\Omega) = 1 \end{cases} \tag{2.31}$$

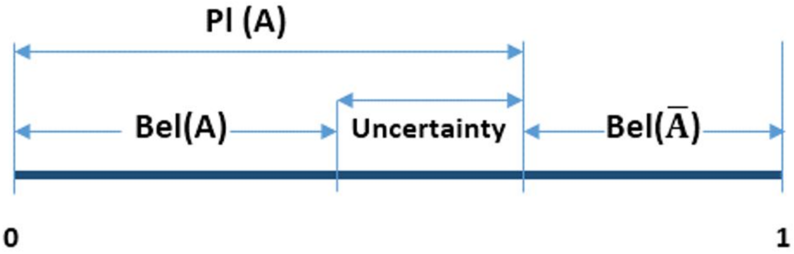


Figure 2.1 – *Uncertainty interval*

	m	Bel	Pl
m		$m(A) = \sum_{B\subset A}(-1)^{\lvert A\rvert+\lvert B\rvert}Bel(B)$	$m(A) = \sum_{B\cap A\neq\emptyset}(-1)^{1+\lvert A\rvert+\lvert B\rvert}Pl(B)$
Bel	$Bel(A) = \sum_{B\subset A} m(B)$		$Bel(A) = 1 - Pl(\overline{A})$
Pl	$Pl(A) = \sum_{B\cap A\neq\emptyset} m(B)$	$Pl(A) = 1 - Bel(\overline{A})$	

Table 2.1 – *Relationship between m, Bel, and Pl*

2.5.3 Fusion process

2.5.3.1 Evidence combination rule

In evidence theory, two BPAs m_1 and m_2 under the same frame of discernment, separately obtained from two independent sources can be combined using Dempster's combination rule which provides a method to compute the orthogonal sum denoted by $m_1 \oplus m_2$ as follows:

$$m_1 \oplus m_2(A) \;=\; \begin{cases} 0 \; if \; A \; = \emptyset \\ \dfrac{1}{1-K} \sum\limits_{A_i \;\cap\; A_j \;=\; A} m_1(A_i)m_2(A_j), \; if \; A \neq \emptyset \end{cases} \tag{2.32}$$

$$K = \sum_{A_i \;\cap\; A_j \;=\; \emptyset} m_1(A_i)m_2(A_j) \tag{2.33}$$

Where K denotes the conflicting factor, quantifying the degree of discordance between two bodies of evidence. It is constrained within the interval $\left[0, 1\right]$.
$1/(1-K)$ is the normalizing factor that ensures the unity property of the fused mass.
The scenario where $K = 0$ indicates perfect consistency and complete agreement among sources, whereas $K = 1$ signifies total conflict among the sources.

Equation 2.32 shows that the fundamental principle underlying Dempster's combination rule involves the direct sum operation applied to the pieces of evidence m_1 and m_2. Therefore, Dempster's combination rule can be alternatively denoted as $m = m_1 \oplus m_2$.

The schematic representation of $m = m_1 \oplus m_2$ is provided in Figure 2.2, where the highlighted pane symbolizes the fused mass associated with hypothesis A.

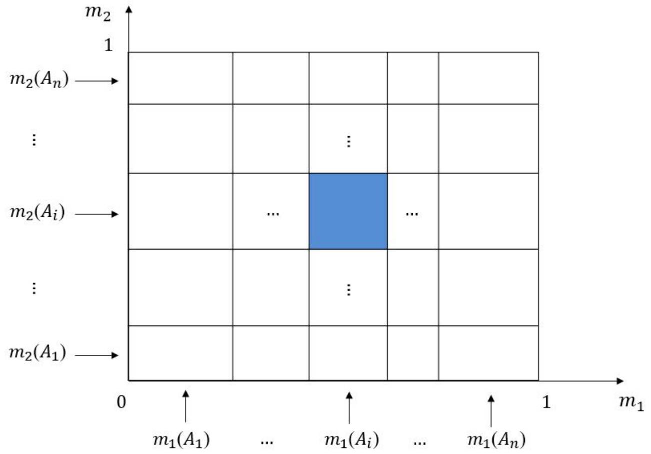


Figure 2.2 – *Dempster's combination rule principle*

Evidently, Dempster's combination rule depicted in Equation 2.32 adheres to both the commutative and associative laws:

$$\begin{cases} m_1 \oplus m_2 = m_2 \oplus m_1 \\ (m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3) \end{cases} \tag{2.34}$$

Accordingly, Dempster's combination rule can be simply extended to the fusion of N body of evidence.

It should be noted that Dempster's combination rule is efficient only upon the absence of conflicts among the pieces of evidence. When the sources are contradictory, applying this rule may lead to unreasonable results, as highlighted in Zadeh's counter-example [73].

2.5.3.2 Decision criteria

After combining the pieces of evidence, a decision is made by selecting an elementary hypothesis among others. This selection process is achieved by maximizing a criterion, where the belief and plausibility of each resulting hypothesis are observed. There are mainly three decision rules: the maximum belief, the maximum plausibility, and the maximum pignistic probability.

- **Maximum of belief:** It consists in choosing the hypothesis that has the greatest credibility value.

$$Dec = arg \max_{A \in \Omega} Bel(A) \tag{2.35}$$

This method is more selective and it is considered as a pessimistic method because Belief function is interpreted as the lower bound of probability.

- **Maximum of Plausibility:** The decision based on maximum plausibility consists in choosing the hypothesis that has the greatest Plausibility value.

$$Dec = \ arg \max_{A \in \Omega} \ Pl(A) \tag{2.36}$$

This method is less selective and it is considered as an optimistic method because Plausibility function is interpreted as the upper bound of probability.

- **Maximum of Pignistic probability:** This rule involves converting the mass function m(A) into a BetP probability function. This conversion is known as the pignistic transformation and can be expressed by the equation:

$$Dec = arg \max_{A \in \Omega} \ BetP(A) \tag{2.37}$$

With,

$$BetP(A) = \sum_{B \in 2^{\Omega}, B \in A} \frac{m(B)}{|B|(1 - m(\emptyset))} \tag{2.38}$$

During this transformation, the belief mass m(A) is evenly distributed across all elements of A. The criterion for selecting an element is based on the highest pignistic probability. This method is considered as compromise between the two previously mentioned methods.

2.6 Comparative analysis

This section is dedicated to conducting a comparative analysis of the various data fusion methods outlined, namely the Bayesian approach, fuzzy logic, and Dempster-Shafer theory of evidence. The objective is to identify their strengths and weaknesses, ultimately selecting the most appropriate formalism for implementation in IoT environments.

2.6.1 Advantages and limitations

2.6.1.1 Bayesian theory

Bayesian theory, grounded in strong mathematical foundations, offers a diverse set of techniques for knowledge representation and a wide range of decision criteria. It effectively handles uncertainty by assigning a measure of certainty to manipulated elements and extrapolating results from random experiments to the entire population. However, it does not easily allow for the representation of imprecision in scenarios where uncertainty is inherently fuzzy or vague and frequently results in a confusion of these two concepts (uncertainty and imprecision). Additionally, Bayesian inference requires high memory resources as it relies on prior knowledge and necessitates simplifying assumptions to ensure manageable and computationally feasible inference.

Bayesian theory establishes a clear and rigorous foundation for data fusion in a multi-source application. Nonetheless, it fails to address conflicts between the sources. Additionally, it encounters challenges in accurately representing and managing partial knowledge or total ignorance situations, the latter is represented by equiprobability, lacking a robust framework for such complexities, which may lead to inconsistent fusion results.

2.6.1.2 Fuzzy logic

Fuzzy set theory serves as a powerful tool for representing and manipulating imprecise data through membership functions. It allows for the incorporation of expert knowledge and subjective assessments into the fusion process and provides a flexible framework for representing and reasoning with linguistic variables and fuzzy relationships, allowing for more intuitive and human-like decision-making processes.

The drawback of fuzzy sets is that they primarily represent the imprecise nature of information, with uncertainty being implicitly represented and only accessible through deduction from various membership functions. Moreover, implementing and optimizing fuzzy logic systems for data fusion tasks can be computationally intensive, especially for large-scale or real-time applications, which may pose challenges in terms of computational efficiency and scalability.

2.6.1.3 Dempster Shafer theory

Dempster-Shafer theory (DST) provides a highly flexible and rich modeling of imperfect knowledge, particularly of imprecision and uncertainty, but also of inconsistency, ambiguity, and incompleteness, by analyzing the capabilities of each source to provide information on each possible decision. Additionally, DST can explicitly model and reason about conflicting evidence. This is especially useful when multiple sources provide contradictory information, allowing for a more nuanced analysis. However, DS theory may encounter challenges when dealing with highly conflicting data, potentially leading to the risk of incorrect fusion.

The theory of belief functions offers a representation of partial knowledge, ranging from total ignorance to perfect knowledge. Belief functions are defined over all subsets of the frame of discernment, unlike in probability theory where they are defined only over singletons, which only measure the probability of belonging to a given hypothesis. Probability theory then becomes a special case of this theory when mass functions are only assigned to singleton hypotheses and in the absence of ignorance $(Bel(A) = Pl(A))$. However, modeling belief functions is a crucial problem in DST for which no generic method exists. Depending on the type of application encountered, there are different methods for developing mass functions. The simplest and most commonly used method is still human expertise, in which coefficients are established manually. Another disadvantage of DST is its computational complexity exponentially growing with the size of the frame of discernment(the number of hypotheses). This complexity can pose challenges for applying DST in real-time applications, therefore, it's crucial to consider optimization techniques like approximation methods or parallel computing to ensure that calculations can be efficiently performed within the required time constraints.

The salient pros and cons of the different data fusion techniques are summarized in Table 2.2.

Approaches	Advantages	Limitations
Bayesian approach	• Less complex with strong mathematical foundations; • Wide range of tools for data modeling and estimation; • Diverse set of decision criteria; • Well-established approach for handling data uncertainty.	• Partial knowledge and total ignorance situations cannot be represented; • Requires simplifying assumptions; • High memory requirements; • High computing time.
Fuzzy logic	• Manipulation of knowledge in natural language using linguistic variables; • Handling imprecise inputs efficiently; • Better results when combined with other techniques; • Fast response (low computing time).	• Higher costs and more computational efforts; • Insufficient flexibility to incorporate prior knowledge; • Challenges in defining IF-THEN rules and membership functions.
Dempster-Shafer theory	• Modeling on a very large space (power set); • Rich and flexible modeling of data imperfections, partial and total ignorance and partial conflict; • Performs very well under uncertainty.	• Growing complexity with the size of the frame of discernment; • Powerless in strong conflict situations; • Challenges in defining mass functions.

Table 2.2 – *Advantages and limitations*

2.6.2 Fusion formalism seclection

In a data fusion scenario, the foremost task is to carefully select a formalism that most closely aligns with the complexities of the specific problem at hand, with the ultimate goal of attaining optimal results. The decision regarding which fusion formalism to use in IoT environments depends mostly on the careful consideration of how well each method addresses data imperfections such as uncertainty, imprecision, and conflict. Each fusion approach, whether it be Bayesian inference, Dempster-Shafer theory, or fuzzy logic, offers unique strengths and weaknesses in addressing different aspects of data imperfection. Bayesian inference provides a probabilistic framework for incorporating uncertainty, while Dempster-Shafer theory excels in handling conflicting evidence. On the other hand, fuzzy logic offers a flexible framework for dealing with imprecision and ambiguity.

Table 2.3 illustrates how different fusion methods address data imperfections such as uncertainty, imprecision, and conflict.

Methods	Data imperfections				
	Uncertainty		Imprecision	Conflict	
	Random	Epistemic		Low	High
Bayesian method	++	+	-	-	-
Fuzzy logic	+	+	++	-	-
DS theory	++	++	++	+	-

Table 2.3 – *Consideration of data imperfections by the different methods*

The Bayesian method manages both epistemic and random uncertainties using probability distributions, making it particularly adept at handling random uncertainties. However, it struggles with imprecise and conflicting data as it requires precise probability distributions.

Fuzzy logic implicitly addresses both epistemic and random uncertainties through deductions from various membership functions. It excels in handling imprecision by allowing partial membership but is not well suited for managing conflicting data.

Dempster-Shafer theory effectively handles both epistemic and random uncertainties and manages imprecision well through belief functions. It excels in dealing with low-conflict situations via its combination rule.

Despite the differences, the mathematical objects used by these theories are similar, though their representational and reasoning power varies. Dempster-Shafer theory stands out by managing imprecision, uncertainty, and particularly low conflict between data sources, making it effective for dealing with most forms of data imperfection commonly encountered in IoT environments, and combining sources with varying degrees of reliability.

2.7 Fuse paradoxes in Dempster-Shafer theory and existing solutions

Dempster–Shafer theory of evidence stands as a robust and flexible mathematical tool for modeling and merging uncertain, imprecise, and incomplete data, and it is widely used in multisensor data fusion applications. However, the combination of contradictory data has always been challenging in DS theory, unreasonable results may arise when dealing with highly conflicting sources. In this section, fuse paradoxes related to Dempster's combination rule are presented along with the existing solutions and alternatives proposed in the literature.

2.7.1 Fuse paradoxes of Dempster's combination rule

Due to the fuzziness and uncertainty in multi-sensor systems, the application of Dempster's combination rule generates counterintuitive results when the information sources are highly conflicting. These conflicting scenarios are referred to as "fuse paradoxes".

Based on the literature, fuse paradoxes primarily arise from the normalization step of the Dempster's combination rule. In Equation 2.32, the variable K represents the level of conflict between evidence m_1 and m_2. As the value of K increases, the conflicts between m_1 and m_2 intensify, leading to counterintuitive results. Notably, when K is equal to 1, Equation 2.32 cannot be applied because the denominator becomes 0 in the normalized factor $1/(1-K)$. There are mainly four types of paradoxes; Completely Conflicting Paradox, 0 trust Paradox, 1 trust Paradox, and Highly conflicting Paradox.

2.7.1.1 Completely Conflicting Paradox

It refers to the situation where two sensors have conflicting outputs, with the output of one sensor completely contradicting that of the other sensor.

Example 1. In the multi-sensor system, there are two pieces of evidence on the frame of discernement $\Omega = \{A, B, C\}$, and that proposition A is true.

Sensor 1: $m_1(A) = 1; m_1(B) = 0; m_1(C) = 0$

Sensor 2: $m_2(A) = 0; m_2(B) = 1; m_2(C) = 0$

In this scenario, the conflicting factor K computed by Equation 2.33 is 1, which reports that the pieces of evidence from sensor 1 and sensor 2 are completely conflicting. The denominator of the equation (combination rule) becomes 0, Dempster's combination rule cannot be applied in this case.

2.7.1.2 0 trust Paradox or One ballot veto paradox

In a multi-sensor system involving more than two sensors, there exists a scenario where one sensor's outputs contradict all the outputs of the other sensors.

Example 2. In the multi-sensor system, assume four pieces of evidence on the frame of discernement $\Omega = \{A, B, C\}$, and that proposition A is true.

Sensor 1: $m_1(A) = 0.5; m_1(B) \ = 0.2; m_1(C) \ = 0.3$

Sensor 2: $m_2(A) \ = 0; m_2(B) \ = 0.9; m_2(C) \ = 0.1$

Sensor 3: $m_3(A) \ = 0.5; m_3(B) \ = 0.2; m_3(C) \ = 0.3$

Sensor 4: $m_4(A) \ = 0; m_4(B) \ = 0.1; m_4(C) \ = 0.9$

By combining these pieces of evidence using Dempster's combination rule, we obtain :

$m(A) \ = 0, \ m(B) = 0.727, \ m(C) = 0.273, \ K = 0.99$

It can be observed that because evidence m_2 completely denies proposition A, the Basic Probability Assignment (BPA) for proposition A in the combined results will always be zero, regardless of how strongly the evidence m_1, m_3, and m_4 support proposition A. This highlights a disadvantage of the Dempster's combination rule, where a single piece of contradictory evidence can act as a veto.

2.7.1.3 1 Trust Paradox or Zadeh paradox

In this case, there is a significant contradiction between one sensor and another sensor, but both sensors share a common focal element with low supporting evidence.

Example 3. In the multi-sensor system, there are two pieces of evidence on the frame of discernement $\Omega = \{A, B, C\}$

Sensor 1: $m_1(A) \ = 0.99 \ ; m_1(B) \ = 0.01 \ ; m_1(C) \ = 0$

Sensor 2: $m_2(A) \ = 0 \ ; m_2(B) \ = 0.01 \ ; m_2(C) \ = 0.99$

By combining these two pieces of evidence with Dempster's combination rule, we obtain :

$m(A) \ = 0, \ m(B) = 1, \ m(C) = 0$

Dempster's combination rule generates counterintuitive results, the wrong proposition B is identified as true with total confidence, even though both pieces of evidence provide minimal support for its validity. In contrast, the possibility of proposition of A and C which were strongly supported by at least one of the pieces of evidence, are completely denied after the combination. This highlights the ineffectiveness of Dempster's combination rule in such scenarios

2.7.1.4 Highly conflicting Paradox

In this scenario, there are multiple sensors with conflicting outputs and different distributions. **Example 4**. In a multi-sensor system, there are five pieces of evidence on the frame of discernement $\Omega = \{A, B, C\}$

Sensor 1: $m_1(A) = 0.3$; $m_1(B) = 0.2$; $m_1(C) = 0.1$; $m_1(\Omega) = 0.4$

Sensor 2: $m_2(A) = 0$; $m_2(B) = 0.9$; $m_2(C) = 0.1$; $m_2(\{\Omega) = 0$

Sensor 3: $m_3(A) = 0.6$; $m_3(B) = 0.1$; $m_3(C) = 0.1$; $m_3(\{\Omega) = 0.2$

Sensor 4: $m_4(A) = 0.7$; $m_4(B) = 0.1$; $m_4(C) = 0.1$; $m_4(\Omega) = 0.1$

Sensor 5: $m_5(A) = 0.7$; $m_5(B) = 0.1$; $m_5(C) = 0.1$; $m_5(\{\Omega) = 0.1$

By applying Dempster's combination rule, the fusion results and the factor of conflict are as follows:

$m(A) = 0; m(B) = 0.9153; m(C) = 0.0847; m(\Omega) = 0; K = 0.8000.$

The fusion results from Dempster's rule assign a higher degree of belief to proposition C, while completely dismissing the accurate proposition A, which is counterintuitive.

2.7.2 Paradox elimination in Dempster's combination rule

To address the fuse paradoxes in DS theory and achieve more reliable combined results, the literature offers several alternatives, which can be broadly classified into three primary categories: (i) Modifying the classical Dempster's combination rule, (ii) Revising the original evidence model before combination and (iii) Hybrid methods, which involve both modifying the evidence model and adjusting Dempster's combination rule.

2.7.2.1 Category 1: Alternatives to the classical Dempster's combination rule

Scholars believe that the unreasonable results generated by Dempster's combination rule in high conflicting situations come from the direct normalization of the conflicting evidence, so they have proposed new combination rules.

Yager [74] proposed transforming conflicting data into total ignorance assigned to the universal set of the frame of discernment denoted by $m(\Omega)$ since conflicting data do not offer useful information:

$$\begin{cases} m(A) = \sum\limits_{A_i \cap A_j = A} m_1(A_i)m_2(A_j) \\ m(\Omega) = 1 - \sum\limits_{A \in 2^\Omega} \sum\limits_{A_i \cap A_j = A} m_1(A_i)m_2(A_j) \end{cases} \tag{2.39}$$

On the contrary, Dubois & Prade's rule [75] regards the data sources as potentially unreliable. They presume that in the event of a conflict between two data sources, at least one of them is reliable. However, given the impossibility of determining the reliable source, they opt to redistribute the conflicting mass across the union of the two sources:

$$m(A) = \sum_{A_i \cap A_j = A} m_1(A_i)m_2(A_j) + \sum_{\substack{A_i \cup A_j = A \\ A_i \cap A_j = \emptyset}} m_1(A_i)m_2(A_j) \tag{2.40}$$

Smet's conjunctive rule [76], also known as the un-normalized Dempster's combination rule, is particularly applicable for open-world assumptions, wherein all the sources are deemed reliable. Smet's concept of belief centers on the understanding that conflict emerges from the incompleteness of the frame of discernment Ω. Consequently, he considers conflict as a form of information allocated to the empty set, treated as an unknown proposition:

$$\begin{cases} m(A) &= \sum\limits_{A_i \cap A_j = A \neq \emptyset} m_1(A_i)m_2(A_j) \\ m(\emptyset) &= \sum\limits_{A \in 2^\Omega} \sum\limits_{A_i \cap A_j = \emptyset} m_1(A_i)m_2(A_j) \end{cases} \tag{2.41}$$

Smet's disjunctive rule [77] considers the union of the pieces of evidence rather than their intersection. In this rule, conflict cannot arise. However, the focal elements of the resulting mass function are expanded, resulting in a loss of specificity:

$\forall\ A \subseteq \Omega$

$$m_1 \cup m_2(A) = \sum_{A_i \cup A_j = A} m_1(A_i) \cup m_2(B_j) \tag{2.42}$$

Lefevre et al. [78] opted for redistributing the conflict proportionally among the focal element sets of all the pieces of evidence.

The previously mentioned rules partially address the conflict issue; however, they do not retain the commutativity and associativity properties satisfied by the classical Dempster's combination rule. Additionally, if counterintuitive results are caused by sensor failure, implementing such solutions would be ineffective.

2.7.2.2 Category 2: Revising the original evidence model

For this category, scholars contend that counter-intuitive results in conflict situations arise from the unreliability of the sources rather than any flaws within Dempster's combination rule itself. They have proposed a pre-processing methodology for mass functions without altering the combination rule. The main idea is to revise and reconstruct the evidence model to reduce the impact of the conflicting evidence on the final fusion result. Broadly, two primary perspectives exist regarding the revision of the evidence model. The first involves a weighted average approach, as proposed by Murphy [79] and the second employs a discounting operation.

Weighted average method

In the process of employing the weighted average method, each individual body of evidence needs to undergo preprocessing before being combined using the classical Dempster's rule.

Suppose that $m_1, m_2, \ldots, m_N$ are N BPAs defined over Ω given by : $\Omega = \{A_1, A_2, ..., A_n\}$. Then the preprocessing operation can be depicted as follows:

$$m_w(A_j) = \sum_{i=1}^{N} w_i \times m_i(A_j) \tag{2.43}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{2.44}$$

Where w_i denotes the weight assigned to the respective BPAs, m_i, and m_w stands for the weighted average of N BPAs.

Once the weighted average evidence are obtained, Dempster's combination rule can be applied to fuse m_w iteratively for $N-1$ times to obtain the final fusion result.

In this methodology, the weight represents the credibility or quality of the corresponding body of evidence, defining its significance in the fusion process. This weight may derive from evaluations of conflict, uncertainty degree, or a combination of both factors.

Discounting method

In the discounting method, each body of evidence should initially undergo discounting using its corresponding discounting factor. Subsequently, the discounted pieces of evidence are combined using the classical Dempster's rule.

Let $m_1, m_2, \ldots, m_N$ be N BPAs defined over Ω given by : $\Omega = \{A_1, A_2, ..., A_n\}$.

The preprocessing for the discounting operation can be depicted as follows:

$$m_j^{\alpha}(A_i) = \left\{ \begin{array}{c} \alpha_j m_j(A_i),\ A_i\ \neq \Omega \\ 1 - \alpha_j +\ \alpha_j m_j(A_i),\ \ A_i\ = \Omega \end{array} \right\} \tag{2.45}$$

Where, α_j represents the discounting factor of BPAs m_j, and m_j^{α} denotes the discounted BPA of m_j. Then, the N discounted BPAs are combined using Dempster's rule to obtain the final fusion result.

The discounting factor is commonly defined either as a function of the conflict between a body of evidence and others, or as a function of the inherent uncertainty within a body of evidence itself , or even as a combination of both.

Related works

To address the issue of counterintuitive results generated by Dempster's combination rule, various alternatives based on preprocessing the evidence model, including discounting and weighted averaging methods, have been proposed in the literature.

Murphy [79] proposed simple averaging mass functions and then combining them using the classical Dempster's combination rule. However, this approach raises concerns as it assigns equal weights to all bodies of evidence without considering their correlation. To address this issue, Yong et al. [80] introduced a weighted average combination rule based on the Jousselme distance to measure the conflict degree between the bodies of evidence. Zhang [81] proposed an improved combining method to evaluate the degree of support between the bodies of evidence using cosine theorem. Yu et al. [82] defined a new distance function, supporting distance function, which measures the correlation between the bodies of evidence and is used to determine weighting factors.

Jing [83] employed the generalized Mahalanobis distance, while Li et al. [84] adopted the Minkowski distance to quantify conflict between the bodies of evidence. Yuan et al. [85] extended the work of [80] by considering the uncertainty of evidence using Deng entropy [86]. Ye et al. [87] introduced the Lance distance function, which combines Minkowski and Mahalanobis distances, along with the spectral angle cosine function, to revise the original evidence model. Tang et al. [88] proposed a novel combination approach using a weighted belief entropy derived from Deng entropy. Lin et al. [89] employed the Euclidean distance to quantify the level of dissimilarity between the bodies of evidence and derive the respective weighting factors. An et al. [90] introduced a new weighted combination method that incorporates fuzzy inference to measure the conflict degree between the bodies of evidence and uses Deng entropy for uncertainty measurement. Li et al. [91] improved the evidence combination method by incorporating Hellinger distance into the Dempster-Shafer evidence theory framework to measure conflict degree and proposed a new belief entropy based on Deng entropy for uncertainty quantification. Deng et al. [92] considered both dissimilarity and inconsistency between the bodies of evidence using Hellinger distance and the sine value of the pignistic vector angle to accurately quantify the conflict degree.

Boulkaboul et al. [61] proposed a novel weighted evidence combination method, DFIOT, for IoT data fusion based on Jousselme distance and an improved Deng entropy, considering contextual parameters such as sensor reliability and information lifetime to determine weighting factors. Li et al. [93] proposed an improved Jousselme distance to evaluate evidence reliability and Tssalis entropy to measure the uncertainty degree for determining weighting factors.

Sun et al. [94] used the Pignistic probability distance function and Deng entropy to revise the original evidence. Yan et al. [95] designed a new belief entropy based on Deng entropy and Zhou's proposed entropy [96] to determine the weight of each BPA. Chen et al. [97] presented an improvement of Dempster's combination rule, considering evidence distance (Jousselme distance), evidence angle, and an improved entropy function.

Ghosh et al. [98] proposed a new method to handle conflict between bodies of evidence using Euclidean distance and a weighted Deng entropy from [99]. Xiao et al. [100]

introduced the evidential correlation coefficient (ECC) inspired by Jiang et al. work [101]to measure the correlation between the pieces of evidence and quantify the conflict degree between them.

Zhu et al. [102] introduced the belief Hellinger distance as a means to assess the discrepancy between the bodies of evidence. This method builds upon the Hellinger distance and takes into account the size of focal elements. Ullah et al. [103] introduced a novel belief entropy based on Deng entropy, which takes into account redundant information in the body of evidence to effectively measure the uncertainty degree.

Zhao et al. [104] presented a new distribution measurement method based on the squared mean of entropy of the BPAs to quantify the conflict degree between the bodies of evidence. Wang et al. [105] introduced a new evidence weight assignment formula by combining Wasserstein distance to compute clarity and credibility degrees of evidence based on the Jousselme distance and Sort-Factor. Ma et al. [106] defined a novel conflict measurement called essential conflict, to determine the weighting factors.

2.7.2.3 Category 3: Hybrid methods

Comparing the above mentioned methods of conflict resolution, the underlying logic of each method becomes evident. The first approach involves eliminating the normalization step in Dempster's combination rule and redistributing the conflict using various measures. On the other hand, the second approach takes into account the fundamental distinctions between propositions from each source in multi-source systems and resolves the conflict by modifying the original evidence. When these two methods are combined, the inherent paradoxes of the DS rule can be effectively addressed.

Building on this idea, Lin et al. [89] proposed a novel combination rule, which involved using Mahalanobis distance for correcting the mass functions during the preprocessing phase. For the combination part, they utilized the Lefevre rule [107], which allows the reallocation of conflicted mass functions to the frame of discernment. Fang et al. [108], introduced Matusita distance function and closeness degree function, to quantitatively assess both the reliability and consistency among the different pieces of evidence, allowing for a comprehensive revision of potentially conflicting evidence. Then instead of directly applying the Dempster's combination rule, an innovative weighted conflict assignment is raised according to the corrected pieces of evidence. This conflict redistribution strategy mitigates the conflicts aising from straightforward normalization. However, these methods lose the commutative and associative properties of Dempster's combination rule.

2.7.2.4 Conflict and uncertainty measures for evidence model revision

Generally, the reliability of a body of evidence depends on two crucial factors: the conflict among different bodies of evidence and the inherent uncertainty within each individual piece of evidence.Consequently, measures of uncertainty and conflict play integral roles in revising the original evidence model within both weighting and discounting methodologies.

Uncertainty measure

In the Dempster-Shafer theory framework, several measures of uncertainty are used to quantify the lack of information or confusion among different sources of information. One of the most common measures of uncertainty is the entropy.

The concept of entropy was first introduced in Physics to describe the level of disorder and chaos in a molecular state within thermodynamics. In information theory, Shannon entropy [109] serves as a metric for uncertainty or information content in a random variable or probability distribution. It measures the average amount of information related to the potential outcomes of a random variable. For a source, which is a discrete random variable X with n symbols, where each symbol x_i has a probability P_i of occurrence, Shannon entropy E_s of the source X is defined as follows:

$$E_s = -\sum_{i=1}^{n} P_i log\ P_i \tag{2.46}$$

Shannon entropy serves as a valuable tool for addressing system uncertainty. While its original purpose lay in quantifying the uncertain information content in information theory. Nevertheless, when it comes to quantifying uncertainty and information volume within the context of Dempster–Shafer evidence theory, the application of Shannon entropy encounters limitations. This is because a mass function, constituting a generalized form of probability, is allocated over the power set of the frame of discernment within Dempster–Shafer evidence theory. Hence, additional entropies have been proposed in the literature to measure uncertainty within the Dempster–Shafer evidence theory framework, with some of them introduced in the following sections.

Deng entropy

Deng entropy is a belief entropy, proposed by Deng [86] as a generalization of Shannon entropy, defined under Dempster–Shafer framework, it takes into consideration the belief mass (BPA - Basic Probability Assignment) of a hypothesis and the cardinality of the elements of the BPA. It serves as an efficient tool to quantify the degree of uncertainty or the volume of information contained in each body of evidence. It is defined as:

$$E_d = -\sum_i m(A_i) log \frac{m(A_i)}{2^{|A_i|} - 1} \tag{2.47}$$

Where A_i represents a hypothesis of a belief function m and $|A_i|$ is the cardinality of the set A_i.

Deng entropy definitely degenerates to Shannon entropy when the mass function is only allocated to singletons (single elements), as follows:

$$E_s = -\sum_i m(A_i) log\ m(A_i) \tag{2.48}$$

Yager's Dissonance Measure

Yager proposed a dissonance measure [110], denoted as E_Y , it is defined as follows:

$$E_Y = -\sum_{A_i \subseteq 2^\Omega} m(A_i) log Pl(A_i) \tag{2.49}$$

Where $Pl(A_i)$ is the plausibility function and $m(A_i)$ is the mass function of proposition A_i .

Weighted Hartley Entropy

Dubois and Prade [111] introduced the Weighted Hartley entropy to quantify uncertainty, it is denoted as E_{DP}, and defined as follows:

$$E_{DP} = -\sum_{A_i \subseteq 2^\Omega} m(A_i) log|A_i| \tag{2.50}$$

Discord measure

Klir and Ramer [112] used the intersection of focal elements within the Frame of Discernment (FOD) to define a discord measure, D_{KR}. It is outlined as follows:

$$D_{KR}(m) = -\sum_{A \subseteq \Omega} m(A) log \sum_{B \subseteq \Omega} m(B) \frac{|A \cap B|}{|B|} \tag{2.51}$$

Zhou et al.'s entropy

Zhou et al. [96] considered the scale of FOD, and defined another belief entropy as follows:

$$E_{Id} = -\sum_{A_i \subseteq \Omega} m(A_i) log \frac{m(A_i)}{2^{|A_i|}-1} e^{\frac{|A_i|-1}{|\Omega|}} \tag{2.52}$$

Where $|A_i|$ represents the number of proposition A_i, and $|\Omega|$ represents the cardinality of Ω, which is the FOD.

Li's new belief entropy

Li et al.[91] defined a new belief entropy to address the limitations related to uncertainty measures. It is expressed as:

$$E_X = -\sum_{A_i \subseteq \Omega} m(A_i) log \frac{m(A_i)}{2^{|A_i|}-1} e^{\frac{|A_i|-1}{2^{|X|}-1}} \tag{2.53}$$

Where X is called the core, it is the union of the focal elements in a body of evidence and $|X|$ is the cardinality of X. If the element A_i is composed of singletons, the new belief entropy degenerates into Shannon entropy.

Yan's new belief entropy

Yan [95] proposed an improved belief entropy based on Deng entropy and Zhou et al.'s belief entropy [96]. The proposed entropy fully considers the correlation between the mass function of a singleton(individual subset) and the mass function of a multi-element subset to quantify the uncertainty within the BOE. It is defined as follows:

$$H_n = -\sum_{A \subseteq \Omega} m(A_i) log \frac{m(A_i) + bel(A_i)}{2\left(2^{|A_i|-1}\right)} \frac{|A_i|-1}{|C|} \tag{2.54}$$

Where $bel(A_i)$ is belief function of A_i, $|A_i|$ is the cardinality of the focal element A, C denotes the cardinality of $BOEs$

Weighted Deng entropy

Chen [97] proposed an improved entropy function to address the inability of Deng entropy to effectively measure the difference among distinct $BOEs$ that are allocated the same mass value. The improved belief entropy is defined as follows:

$$E_w = -\sum_{A \subseteq \Omega} m(A_i) log \frac{m(A_i)}{2\left(2^{|A_i|}-1\right)} \frac{|A_i|}{|C|} \tag{2.55}$$

Where $|A_i|$ denotes the cardinality of the focal element A_i , $|C|$ is the total number of focal elements in this BOE, and $\frac{|A_i|}{|C|}$ represents the uncertain information in a BOE that has not been considered by Deng entropy.

Tang's entropy

Tang [99] proposed a weighted belief entropy that addresses more uncertain information in the *BOE* by including the scale of the *FOD*, denoted as $|\Omega|$, and the relative scale of a focal element with respect to the FOD, denoted as $\frac{|A_i|}{|\Omega|}$. The new belief entropy E_{wd} is defined as follows:

$$E_{wd} = -\sum_i \frac{|A_i|.m(A_i)}{|\Omega|} log \frac{m(A_i)}{2\left(2^{|A_i|}-1\right)} \tag{2.56}$$

Ullah's improved belief entropy

Ullah [103] proposed an improved belief entropy based on Deng entropy, which considers the available redundant information in the body of evidence *BOE*, to improve the accuracy of the uncertainty measure. The proposed entropy E_p is defined as follows:

$$E_p = -\sum_i m\left(A^{'}\right) log \frac{m\left(A^{'}\right)}{2\left(2^{|A^{'}|}-1\right)} \tag{2.57}$$

$$A^{'} = m(A_i) \cup m(A_j) \tag{2.58}$$

Where A' represents the union of the two *BOEs* m_1 and m_2.

Zhao's modified entropy

Zhao et al [113] proposed an improved belief entropy to measure uncertainty of the bodies of evidence based on Deng entropy and the belief interval, composed of the belief function as the lower bound and the plausibility function as the upper bound. This method degenerates to Shannon entropy when the basic probability assignments are Bayesian.

$$E = -\sum_{A \subseteq \Omega} m(A_i) log \frac{m(A_i)}{2^{|A_i|}-1} e^{\sum_{A_j \subseteq \Omega,\, A_j \neq A_i} \frac{|A_i \cap A_j|}{2^{|A_i|}}} \tag{2.59}$$

Mambe et al.'s modifid entropy

Inspired by Deng and Zhou entropies, Mambe et al [114] defined a new measure that takes into account the number of elements of all parts of the FOD represented by $2^{|\ \Omega|}$, which is not perceived in Deng and Zhou et al. entropies. The modified entropy is given as follows:

$$E_{Nm} = -\sum_{A \subseteq \Omega} m(A_i) log \frac{m(A_i)}{2^{|A_i|}-1} e^{\frac{|A_i|-1}{2^{|\Omega|}}} \tag{2.60}$$

Conflict measure

Conflict measure in the context of Dempster-Shafer theory quantifies the degree of inconsistency or discrepancy between pieces of evidence. It provides a way to assess the disagreement or discordance among different sources of information, which is crucial for making informed decisions in uncertain environments. Among the diverse array of conflict measurement methods, distance measures are some of the most frequently employed techniques. These measures quantify the dissimilarity between evidence sources by assessing their geometric or probabilistic distances. Additionally, consensus measures offer valuable insights into the level of agreement among evidence sources. Some commonly used conflict measures are listed below.

Euclidean distance

The Euclidean distance is also used to measure the conflict degree between the bodies of evidence within the Demspter-Shafer theory framework, it is defined as follows:

$$d(m_i, m_{avg}) = \sqrt{\sum_{j=1}^{n} \left[m_i(A_j) - m_{avg}(A_j)\right]^2} \tag{2.61}$$

Jousselme distance

Jousselme distance is the most frequently used evidence distance, it was proposed by Jousselme et al. [115], its principle comes from Cuzzolin's geometric interpretation of evidence theory [116], where the frame of discernement is considered to be a 2^N linear space. The Jousselme distance is defined as:

$$d_J(m_i, m_j) = \sqrt{\frac{1}{2}\left(\overrightarrow{m_i} - \overrightarrow{m_j}\right)^T D \left(\overrightarrow{m_i} - \overrightarrow{m_j}\right)} \tag{2.62}$$

With D $2^n \times 2^n$ matrix known as Jaccard matrix whose elements are defined as:

$$D(A_i, A_j) = \frac{|A_i \cap A_j|}{|A_i \cup A_j|} \quad A_i,\ A_j\ \in\ 2^{\Omega} \tag{2.63}$$

Hellinger distance

Hellinger distance for D– S evidence theory Hellinger distance is a complete distance metric defined in the probability distribution space; it is considered as the probabilistic analog of Euclidean Distance. It was expressed in terms of the Hellinger integral initiated by Hellinger in 1909. Hellinger distance is very stable and reliable and it is widely used to measure the dissimilarity of two probability distributions and it can be applied to evidence theory, to measure the dissimilarity between two pieces of evidence. In a finite complete frame of discernment, Hellinger distance between two bodies of evidence is defined as:

$$d_H(m_1, m_2) = \frac{1}{2} \sum_{i=1}^{n} \left\| \sqrt{m_1(A_i)} - \sqrt{m_2(A_i)} \right\|_2 \tag{2.64}$$

Mahalanobis distance

The Mahalanobis distance, introduced by the Indian statistician P. C. Mahalanobis, offers a means of determining the distance between two points by considering covariance. This method enables the computation of the most concise distance from a sample to the "center of gravity" within the sample set or the assessment of the similarity between two unfamiliar sample sets. It was introduced into the framework of Dempster Shafer theory [117] to calculate the distance between the bodies of evidence as follows:

$$GD_m(m_i, m_j) = \sqrt{\frac{1}{2} \left(\overrightarrow{m_i} - \overrightarrow{m_j} \right)^T P^+ \left(\overrightarrow{m_i} - \overrightarrow{m_j} \right)} \tag{2.65}$$

Where P stands for the covariance matrix between the bodies of evidence.

2.7.2.5 Lance distance

Lance distance [87] is the average distance of evidence m_i, m_j among different hypotheses in the power set 2^Ω, it is defined as follows:

$$d_{ij}(L) = \frac{1}{n} \sum_{1}^{n} \frac{|m_i - m_j|}{\left(m_i + m_j \right)} \tag{2.66}$$

Betting commitement distance

Tessem [118] proposed the betting commitement distance based on the Pignistic probability transformation, to measure the conflict degree among the bodies of evidence.
Let m be a BPA on Ω, the associated Pignistic probability transformation $Betm_i$: $\Omega \to \left[0, 1\right]$ is given by:

$$BetP_m = \sum_{A \subseteq \Omega} \frac{1}{|A_i|} \frac{m(A_i)}{1 - m(\emptyset)}, m(\emptyset) \neq 1 \tag{2.67}$$

The betting commitement distance between m_i and m_j is then defined as follows:

$$DifBet(m_i, m_j) = max_{A_i \subseteq \Omega} \left(|Betm_i(A_i) - Betm_j(A_i)| \right) \tag{2.68}$$

Cosine value

Cosine value is an effective tool to evaluate the similarity between two evidence geometrically, by taking the cosine of the angle the vectors of the two evidence make in their dot product. In a finite frame of discernement Ω, the similarity between two BPAs using cosine value is defined by:

$$cos(\theta) = \frac{\overrightarrow{m_i}\ .\overrightarrow{m_j}^T}{\left\|\overrightarrow{m_i}\right\|_2 \left\|\overrightarrow{m_j}\right\|_2} \tag{2.69}$$

θ: The angle formed by the vectors of the evidence m_i and m_j
$\overrightarrow{m_i}\ .\overrightarrow{m_j}^T$: The inner product of $\overrightarrow{m_i}$ and $\overrightarrow{m_j}$.
$\left\|\overrightarrow{m_i}\right\|_2, \left\|\overrightarrow{m_j}\right\|_2$: Vectors' norm

Song et al.'s correlation coefficient

Song et al [119] defined a correlation coefficient considering the similarity among the subsets of the frame of discernment using Jaccard matrix as follows:

$$C_{sw}(m_i, m_j) = \frac{<\ m_i', m_j' >}{\left\|m_i'\right\| \left\|m_j'\right\|} \tag{2.70}$$

In which $m^{'}$ is expressed as:

$$\begin{cases} m_i^{'} = m_i.D \\ m_j^{'} = m_j.D \end{cases}$$

Where D is Jaccard matrix defined in equation 2.63
The conflict coefficient is then obtained by:

$$K_{sw}(m_i, m_j) = 1 - \ C_{sw}(m_i, m_j) \tag{2.71}$$

Cheng and Xiao's distance

Based on Jousselme distance, Chang and Xiao have introduced a novel distance measure [120] using a newly defined similarity coefficient. In this measure, both of the involved bodies of evidence contribute to measuring the similarity degree as follows:

$$d_{CX}(m_1, m_2) = \sqrt{\frac{1}{2}\left(\overrightarrow{m_1} - \overrightarrow{m_2}\right)^T D_\alpha \left(\overrightarrow{m_1} - \overrightarrow{m_2}\right)} \tag{2.72}$$

Where D_α is $2^n \times 2^n$ matrix with elements:

$$D_\alpha\left(A_i, A_j\right) = \frac{|A_i \cap A_j|}{|A_i|}\frac{|A_i \cap A_j|}{|A_j|} \tag{2.73}$$

2.8 Conclusion

In addressing fusion challenges, the primary task is to choose a framework that best aligns with the specific problem, aiming to achieve optimal results.

Throughout this chapter, we have explored the most commonly used mathematical techniques for data fusion, specifically focusing on decision-level fusion methods that account for various imperfections in IoT data. These include the Bayesian approach, fuzzy logic, and Dempster-Shafer theory. A comparative analysis of these methods has been conducted, demonstrating that each approach offers distinct modeling and information processing capabilities.

Dempster-Shafer theory emerges as a robust, effective, and flexible option, proving to be the most suitable approach for managing uncertainty and conflict in IoT environments. However, the combination of contradictory data poses an ongoing challenge within DS theory, potentially leading to unreasonable results. The limitations of Dempster-Shafer theory and the various solutions proposed in the literature for managing uncertainty and conflict within the Dempster-Shafer framework have been discussed. In the next chapter, we'll present our proposed strategies to overcome the outlined limitations.

3

The proposed improved evidence combination approaches

Contents

3.1 Introduction

To overcome the classical Dempster's combination rule flaw and fuse highly conflicting evidence without generating counter-intuitive results, enhanced evidence combination approaches are developed and presented in this chapter. They are based on preprocessing the mass functions before the combination. Weights representing the degree of confidence given to data sources are determined using various factors. The methods are specially designed for uncertainty measure and evidence conflict management while taking into consideration the heterogeneity of IoT data, enabling the fusion system to reach effective decision-making.

3.2 An Advanced Weighted Evidence Combination Approach (AWECA)

The novel weighted evidence combination method proposed to handle conflict when combining the bodies of evidence is based on three main tools; evidence distance, evidence angle and belief entropy [121]. Evidence distance represents dissimilarity between the bodies of evidence, whereas evidence angle describes the consistency between them, both measurements are employed to quantify the degree of conflict among the sources and contribute to determining the credibility degree of each source. On the other hand, belief entropy measures the uncertainty level within each body of evidence. Both of these measurements are subsequently employed in constructing the relevant weighting factors, with greater weights assigned to sources that are well supported by the others,and conversely, lower weights to less supported sources, which helps alleviate the conflicting impact on the final fusion results.

The steps for the improved combination method are described in the following:

3.2.1 Conflict measurement

As previously mentioned, both the evidence distance and evidence angle are used to quantify the degree of conflict among the sources.

For the provided mass functions, which correspond to the data collected from various sources, Equation 2.62 is used to calculate the Jousselme distance between each pair of bodies of evidence.

The $N \times N$ distance matrix $D_J(m_i, m_j)$ is expressed below:

$$D_J(m_i, m_j) = \begin{pmatrix} 0 & d_J(m_1, m_2) & \dots & d_J(m_1, m_N) \\ d_J(m_2, m_1) & 0 & \dots & d_J(m_2, m_N) \\ \vdots & \vdots & \vdots & \vdots \\ d_J(m_N, m_1) & d_J(m_N, m_2) & \dots & 0 \end{pmatrix} \tag{3.1}$$

Based on equation 2.69, the cosine value between every pair of bodies of evidence can be obtained. The $N \times N$ cosine value matrix $cos(m_i, m_j)$ is then expressed as follows:

$$cos(m_i, m_j) = \begin{pmatrix} 0 & cos(m_1, m_2) & \dots & cos(m_1, m_N) \\ cos(m_2, m_1) & 0 & \dots & cos(m_2, m_N) \\ \vdots & \vdots & \vdots & \vdots \\ cos(m_N, m_1) & cos(m_N, m_2) & \dots & 0 \end{pmatrix} \tag{3.2}$$

3.2.2 Credibility degree of evidence

The credibility degree is used to estimate the reliability of each source and validate the accuracy of the provided information. Evaluating the reliability is very crucial especially when handling multiple sources that may present conflicting information. This assessment should be conducted before the fusion process to determine the appropriate weights for each piece of evidence.

First, the similarity degree between every pair of bodies of evidence (BOEs) is calculated by combining both the cosine similarity and the similarity derived from the evidence distance, these two measures are mutually complementary. As introduced before, the evidence distance and evidence angle establish that the smaller the distance between two Bodies of Evidence (BOEs), the more similar they are. Conversely, the larger the cosine value of the evidence angle, the more consistent these two BOEs are.

Thus, we formulate the similarity measure, $sim(m_i, m_j)$, between m_i and m_j as follows:

$$sim(m_i, m_j) = \Big(1 - d_J(m_1, m_2)\Big)\, cos(m_i, m_j) \tag{3.3}$$

The $N \times N$ similarity matrix can be represented by:

$$sim(m_i, m_j) = \begin{pmatrix} 1 & sim(m_1, m_2) & \dots & sim(m_1, m_N) \\ sim(m_2, m_1) & 1 & \dots & sim(m_2, m_N) \\ \vdots & \vdots & \vdots & \vdots \\ sim(m_N, m_1) & sim(m_N, m_2) & \dots & 1 \end{pmatrix} \tag{3.4}$$

Next, the support degree of a each body of evidence BOE m_i (i = 1, 2, ..., n) can be calculated using the previously defined similarity measure, as follows:

$$sup(m_i) = \sum_{j=1, j\neq i}^{N} sim(m_i, m_j) \tag{3.5}$$

Finally, we can assess the reliability of each source by calculating the credibility degree of each evidence as follows:

$$CRD(m_i) = \frac{Sup(m_i)}{\sum_{j=1}^{N} sup(m_j)} \tag{3.6}$$

3.2.3 Uncertainty degree of evidence

Several factors related to uncertainty such as noisy, erroneous or missing data, etc.) can hinder the decision-making process. Therefore, it is crucial, in the conflict resolution process, to determine the degree of uncertainty associated with each information source. In this step, we use Deng entropy to obtain a better weighting factor, the measure of each mass function is calculated using Equation 2.47.

Based on Deng Entropy, information volume I_V associated with each body of evidence is calculated as follows:

$$Iv(m_i) = e^{E_d} = e^{-\sum_i m(A_i) log_2 \frac{m(A_i)}{2^{|A_i|}-1}} \tag{3.7}$$

3.2.4 Weighted BPAs calculation

3.2.4.1 Weights'determination

The weights assigned to each body of evidence can be accurately determined based on the volume of information and the credibility degree as follows:

$$w_i = \frac{CRD(m_i) \times Iv(m_i)}{\sum_{j=1}^{N} CRD(m_j) \times Iv(m_j)} \tag{3.8}$$

Hence, the weighted average bodies of evidence BOEs denoted as m_w are obtained by:

$$m_w(A) = \sum_{i=1}^{N} w_i \times m_i(A) \tag{3.9}$$

3.2.5 Evidence combination

In the combining process, the classical Dempster's combination rule is iteratively applied $N-1$ times to combine the weighted bodies of evidence using the following formula:

$$\begin{aligned}&\Big(m_1 \oplus m_2 \oplus \ldots \oplus m_N\Big)\Big(A_i\Big)\\&= \Bigg(\bigg(\Big((m_w(A_i) \oplus m_w(A_i))_{(1)} \oplus m_w(A_i)\Big)_{(2)} \oplus m_w(A_i)\bigg)_{(3)} \oplus \ldots \oplus m_w(A_i)\Bigg)_{(N-1)}\end{aligned} \tag{3.10}$$

The flowchart illustrating the proposed weighted evidence combination approach is provided in Figure 3.1

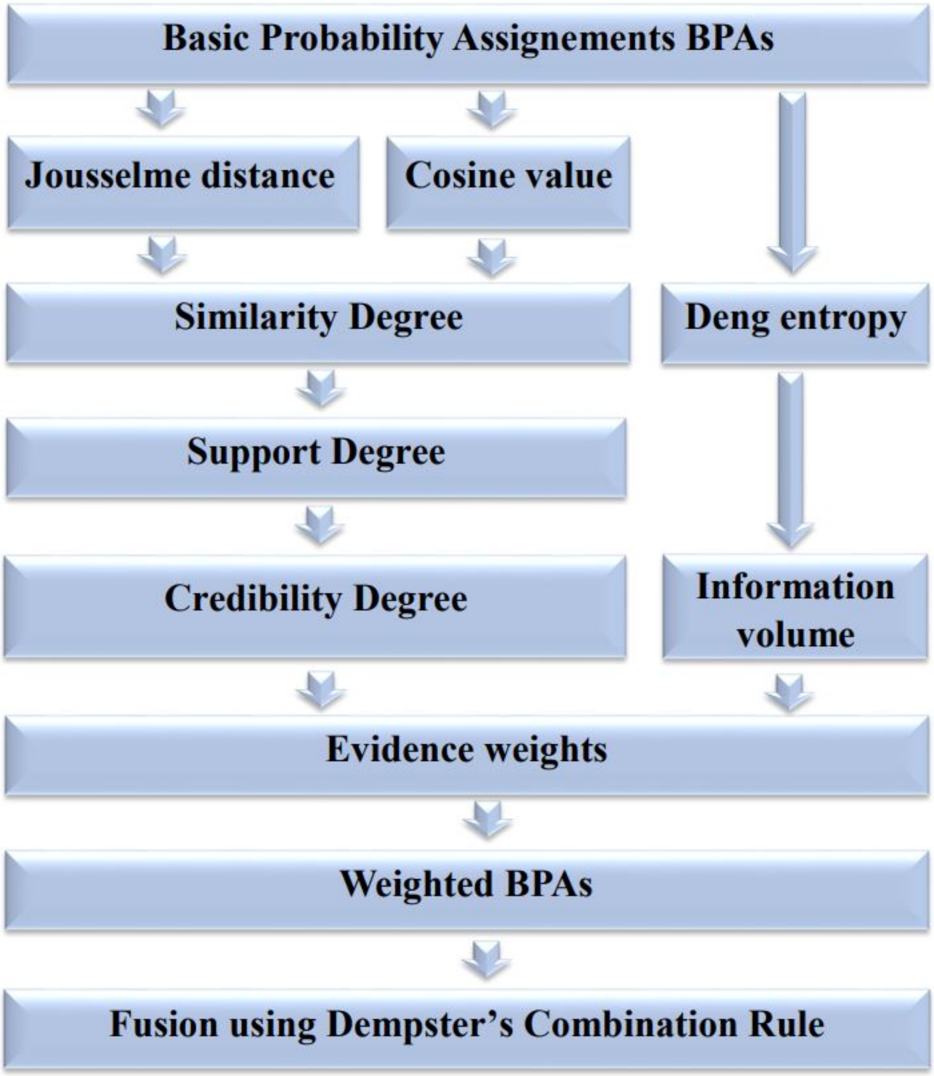


Figure 3.1 – *Flowchart of AWECA method*

3.3 An Improved Evidence Distance-based Combination Approach (IDECA)

In this section, we introduce an enhanced evidence combination method that primarily relies on a developed, improved evidence distance to handle both conflict and uncertainty within IoT environments [122]. Similar to AWECA method, this novel approach evaluates the relative importance of each information source in the fusion process by assigning evidence weights.

The key idea is to allocate high weights to reliable sources and, conversely, low weights to less reliable ones, to mitigate their conflicting influence on the final fusion result. This, in turn, enhances the accuracy of decision-making.

Our first contribution is the definition of an improved evidence distance based on Hellinger distance, which effectively quantifies the degree of conflict between the bodies of evidence. It takes into account the interdependencies between these pieces of evidence through Jaccard matrix, satisfying key metric properties (non-negativity, symmetry, positive definiteness, trigonometric inequality) while providing a more robust measure of conflict.

Subsequently, we introduce a novel evidence fusion strategy built upon the improved evidence distance to address the conflict degree between the bodies of evidence and employs Deng entropy to quantify the uncertainty associated with each body of evidence. To assign weights, we design reward and penalty functions, reliable pieces of evidence are rewarded with heavier weights, amplifying their impact on the final fusion result, while less reliable sources are allotted lower weights, thereby reducing their impact on the final result. Finally, we apply the classical Dempster's rule to combine the modified bodies of evidence.

3.3.1 The improved evidence distance

Drawing inspiration from the concept of belief functions' transformation presented in Song et al.'s work [119], we introduce a novel enhanced evidence distance based on the Hellinger distance for Dempster-Shafer evidence theory. This newly introduced distance metric integrates the correlation among the different bodies of evidence through the utilization of Jaccard matrix, thereby enhancing its effectiveness in measuring the level of conflict among these bodies of evidence.

Our improved evidence distance adheres to the fundamental requirements of a true metric, satisfying properties such as non-negativity, absence of degeneracy, symmetry, and the triangle inequality. It's worth noting that in cases where all the elements are singletons, the mass function conforms to the classical probability distribution, and the Jaccard

matrix corresponds to the identity matrix. Consequently, the improved evidence distance degenerates to the traditional Hellinger distance.
We define the improved evidence distance as follows:

$$d_{IH}(m_1, m_2) = \frac{1}{2}\sum_{i=1}^{n} \left\| \sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)} \right\|_2 \tag{3.11}$$

In which $m^{'}$ is expressed as:

$$\begin{cases} m_1^{'} = m_1.D \\ m_2^{'} = m_2.D \end{cases} \tag{3.12}$$

Where D is Jaccard matrix of size $2^n \times 2^n$ whose elements are defined in equation 2.63 The improved evidence distance satisfies the properties of:

1. Non-Negativity $0 \leqslant d_{IH}(m_1, m_2) \leqslant 1$
2. Symmetry $d_{IH}(m_1, m_2) = d_{IH}(m_2, m_1)$
3. Triangle inequality $d_{IH}(m1, m2) + d_{IH}(m_2, m_3) \ \leqslant \ d_{IH}(m_1, m_3)$
4. Positive definiteness $d_{IH}(m_1, m_2) = 0, \ if \ and \ only \ if \ m_1 \ = \ m_2$

Proofs
In the following, the properties of non-negativity, symmetry, triangle inequality and positive definiteness of the improved distance are verified.
The equation 3.11 can be written as:

$$d_{IH}(m_1, m_2) = \sum_{i=1}^{n} \sqrt{\frac{1}{2}\left(\sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)}\right)^2} \tag{3.13}$$

Proof 1 : Non-Negativity $0 \leqslant d_H(m_1, m_2) \leqslant 1$

$$\begin{aligned} d_{IH}{}^2(m_1, m_2) &= \frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)}\right)^2 \\ &= \frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)}\right)\left(\sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)}\right) \\ &\leqslant \frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)}\right)\left(\sqrt{m_1'(A_i)} + \sqrt{m_2'(A_i)}\right) \\ &= \frac{1}{2}\sum_{i=1}^{n}\left(m_1'(A_i) - \ m_2'(A_i)\right) \end{aligned} \tag{3.14}$$

Since m on Ω satisfies:

$$0 \leqslant m(A_i) \leqslant \ 1 \tag{3.15}$$

And the coefficient $D(A_i, A_j)$ of D has the property:

For all A_i, A_j of 2^{Ω} :

$$0 \leqslant D(A_i, A_j) \leqslant 1 \tag{3.16}$$

We can deduce that

$$0 \leqslant m^{'}(A_i) \leqslant 1 \tag{3.17}$$

Thus

$$0 \leqslant \sum_{i=1}^{n} \left(m_1'(A_i) - m_2'(A_i) \right) \leqslant 2 \tag{3.18}$$

Which implies

$$0 \leqslant d_{IH}(m_1, m_2) \leqslant 1 \tag{3.19}$$

Non negativity property of the proposed evidence distance is proved.

Proof 2 : Symmetry $d_{IH}(m_1, m_2) = d_{IH}(m_2, m_1)$

We have

$$d_{IH}(m_1, m_2) = \sum_{i=1}^{n} \sqrt{\frac{1}{2} \left(\sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)} \right)^2} \tag{3.20}$$

And

$$d_{IH}(m_2, m_1) = \sum_{i=1}^{n} \sqrt{\frac{1}{2} \left(\sqrt{m_2'(A_i)} - \sqrt{m_1'(A_i)} \right)^2} \tag{3.21}$$

It can be noted that:

$$\left(\sqrt{m_2'(A_i)} - \sqrt{m_1'(A_i)} \right)^2 = \left(\sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)} \right)^2 \tag{3.22}$$

Thus

$$d_{IH}(m_1, m_2) = d_{IH}(m_2, m_1)$$

Therefore, the symmetry property of the proposed evidence distance is proved.

Proof 3 : Triangle inequality $d_{IH}(m_1, m_2) + d_{IH}(m_2, m_3) \leqslant d_{IH}(m_1, m_3)$

We have:

$$d_{IH}(m_1, m_2) + d_{IH}(m_2, m_3) =$$

$$\frac{1}{\sqrt{2}} \left[\sum_{i=1}^{n} \left(\sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)} \right)^2 \right]^{\frac{1}{2}} + \frac{1}{\sqrt{2}} \left[\sum_{i=1}^{n} \left(\sqrt{m_2'(A_i)} - \sqrt{m_3'(A_i)} \right)^2 \right]^{\frac{1}{2}} \tag{3.23}$$

We use Minkowski inequality given by:

$$\left[\sum_{i=1}^{n} \left(a_i + b_i \right)^p \right]^{\frac{1}{p}} \leqslant \left[\sum_{i=1}^{n} \left(a_i \right)^p \right]^{\frac{1}{p}} + \left[\sum_{i=1}^{n} \left(b_i \right)^p \right]^{\frac{1}{p}} \tag{3.24}$$

Where $p > 1$ and $a_i, b_i > 0$

We get:

$$
\begin{aligned}
& d_{IH}(m_1, m_2) + d_{IH}(m_2, m_3) \\
& \quad = \frac{1}{\sqrt{2}} \left[\sum_{i=1}^{n} \left(\sqrt{m'_1(A_i)} - \sqrt{m'_2(A_i)} \right)^2 \right]^{\frac{1}{2}} + \frac{1}{\sqrt{2}} \left[\sum_{i=1}^{n} \left(\sqrt{m'_2(A_i)} - \sqrt{m'_3(A_i)} \right)^2 \right]^{\frac{1}{2}} \\
& \quad = \frac{1}{\sqrt{2}} \left\{ \left[\sum_{i=1}^{n} \left(\sqrt{m'_1(A_i)} - \sqrt{m'_2(A_i)} \right)^2 \right]^{\frac{1}{2}} + \left[\sum_{i=1}^{n} \left(\sqrt{m'_2(A_i)} - \sqrt{m'_3(A_i)} \right)^2 \right]^{\frac{1}{2}} \right. \\
& \quad \left. \leqslant \frac{1}{\sqrt{2}} \left[\sum_{i=1}^{n} \left(\sqrt{m'_1(A_i)} - \sqrt{m'_2(A_i)} \right)^2 + \sum_{i=1}^{n} \left(\sqrt{m'_2(A_i)} - \sqrt{m'_3(A_i)} \right)^2 \right]^{\frac{1}{2}} \right\}
\end{aligned}
\tag{3.25}
$$

Since:

$$
\left(\sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)} \right) + \left(\sqrt{m_2'(A_i)} - \sqrt{m_3'(A_i)} \right) \leqslant \left(\sqrt{m_1'(A_i)} - \sqrt{m_3'(A_i)} \right) \tag{3.26}
$$

We can deduce:

$$
d_{IH}(m_1, m_2) + d_{IH}(m_2, m_3) \leqslant \frac{1}{\sqrt{2}} \left[\sum_{i=1}^{n} \left(\sqrt{m_1'(A_i)} - \sqrt{m_3'(A_i)} \right)^2 \right]^{\frac{1}{2}} \tag{3.27}
$$

And:

$$
\frac{1}{\sqrt{2}} \left[\sum_{i=1}^{n} \left(\sqrt{m_1'(A_i)} - \sqrt{m_3'(A_i)} \right)^2 \right]^{\frac{1}{2}} = d_{IH}(m_1, m_3) \tag{3.28}
$$

Therefore:

$$
d_{IH}(m_1, m_2) + d_{IH}(m_2, m_3) \ \leqslant \ d_{IH}(m_1, m_3) \tag{3.29}
$$

Thus, the triangle inequality property of the proposed evidence distance is proved.

Proof 4 : Positive definiteness $d_{IH}(m_1, m_2) = 0$, if and only if $m_1 = m_2$

$$
\begin{aligned}
d_{IH}(m_1, m_2) = 0 \Leftrightarrow & \ \frac{\left\| \sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)} \right\|_2}{\sqrt{2}} = 0 \\
\Leftrightarrow & \ \left\| \sqrt{m_1'(A_i)} - \sqrt{m_2'(A_i)} \right\|_2 = 0 \\
\Leftrightarrow & \ m_1^{'} = m_2^{'} \\
\Leftrightarrow & \ m_1.D = m_2.D \\
\Leftrightarrow & \ m_1 = m_2
\end{aligned}
\tag{3.30}
$$

Finally, the positive definiteness property of the proposed evidence distance is successfully proved.

As demonstrated by the proofs provided above, the improved evidence distance complies with all the necessary criteria, making it a valid metric within the Dempster-Shafer theory framework.

3.3.2 Conflict measure

According to Equation 3.11, the improved evidence distance between every two bodies of evidence $m_i(i = 1, 2, \dots, N)$ and $m_j(j = 1, 2, \dots, N)$ is calculated to measure the conflict degree.

The $N \times N$ distance matrix D_{IH} is expressed below:

$$D_{IH}(m_i, m_j) = \begin{pmatrix} 0 & d_{IH}(m_1, m_2) & \dots & d_{IH}(m_1, m_N) \\ d_{IH}(m_2, m_1) & 0 & \dots & d_{IH}(m_2, m_N) \\ \vdots & \vdots & \vdots & \vdots \\ d_{IH}(m_N, m_1) & d_{IH}(m_N, m_2) & \dots & 0 \end{pmatrix} \tag{3.31}$$

3.3.3 Credibility degree

Initially, the degree of similarity between each pair of evidence is determined using the formula outlined in [91] as follows:

$$sim(m_i, m_j) = \left(1 - \sqrt{d_{IH}(m_i, m_j)}\right) e^{-\sqrt{d_{IH}(m_i, m_j)}} \tag{3.32}$$

The $N \times N$ similarity matrix can be written now as:

$$SIM(m_i, m_j) = \begin{pmatrix} 1 & sim(m_1, m_2) & \dots & sim(m_1, m_N) \\ sim(m_2, m_1) & 1 & \dots & sim(m_2, m_N) \\ \vdots & \vdots & \vdots & \vdots \\ sim(m_N, m_1) & sim(m_N, m_2) & \dots & 1 \end{pmatrix} \tag{3.33}$$

Then the support degree of each evidence can be evaluated using the previously calculated similarity degrees as follows:

$$sup(m_i) = \sum_{j=1, j\neq i}^{N} sim(m_i, m_j) \tag{3.34}$$

Finally, the degree of credibility is computed to represent the level of trustworthiness attributed to each piece of evidence.The higher the credibility, the more reliable the evidence. This is expressed as:

$$CRD(m_i) = \frac{sup(m_i)}{\sum\limits_{j=1, j\neq i}^{N} sup(m_i, m_j)} \tag{3.35}$$

3.3.4 Uncertainty degree

Based on Equation 2.47, Deng entropy for each evidence is calculated to quantify the uncertainty degree.

3.3.5 Weighted BPAs calculation

3.3.5.1 Reliability condition

We establish a condition to evaluate the reliability of each piece of evidence by defining a threshold, denoted by α, that distinguishes between reliable and unreliable evidence. The threshold α is defined as follows:

$$\alpha = \frac{\sum_{i=1}^{N} CRD(m_i)}{N} \tag{3.36}$$

When the credibility of a piece of evidence exceeds the threshold (i.e., $CRD(m_i) \geqslant \alpha$) the source is considered reliable. Conversely, if the credibility falls below the threshold (i.e., $CRD(m_i) < \alpha$), the source is deemed unreliable.

3.3.5.2 Weights' determination

1. **Initial weights**

   We establish the initial weights by evaluating the fulfillment of the reliability condition. The objective is to increase the weights that surpass the threshold while decreasing the weights that fall below it, as outlined below:

   If $CRD(m_i) \geqslant \alpha$, a reward function is defined as follows:

$$w_I(m_i) = e^{E_d(m_i)} \tag{3.37}$$

   If $CRD(m_i) < \alpha$, a penalty function is defined as follows:

$$w_I(m_i) = e^{-\left(E_{dmax} + 1 - E_d(m_i)\right)} \tag{3.38}$$

2. **Final weights**

   The final weights can be determined based on both the credibility degree and the initial weight, resulting in the following definition:

$$w(m_i) = \frac{CRD(m_i) \times w_I(m_i)}{\sum_{j=1}^{N} CRD(m_j) \times w_I(m_j)} \tag{3.39}$$

The final weights obtained are then used to modify the original BPAs as follows:

$$m_w(A_i) = \sum_{j=1}^{N} w(m_j) \times m_j(A_i) \tag{3.40}$$

3.3.6 Evidence combination

Finally, for N body of evidence $m_1, m_2, ..., m_N$, classical Dempster's combination rule is applied $N-1$ times to get the final fusion result of the weighted BPAs.

The decision is made based on the maximum belief, the hypothesis with the strongest support is selected.

The flowchart of the IDECA approach is depicted in Figure 3.2

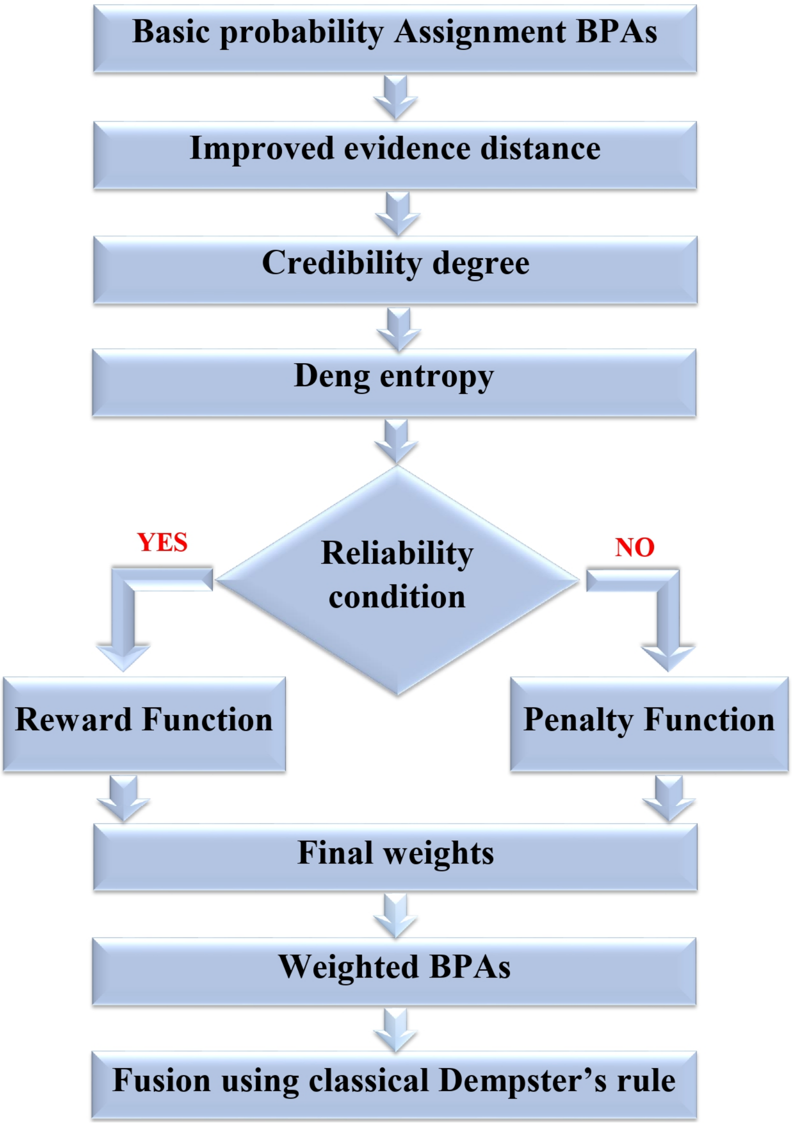


Figure 3.2 – *Flowchart of IDECA*

3.4 Fuzzy Similarity measure-based Evidence Combination Approach (FSECA)

The advanced evidence combination approach based on fuzzy similarity involves preprocessing the evidence model before the combination to address the counterintuitive issues associated with the classical Dempster's rule. The main concept of this method is to integrate the fuzzy inference mechanism into the similarity measure model to effectively quantify the degree of conflict among the pieces of evidence, using the enhanced distance proposed in section 3.3.1 and cosine value. Expanding on this, a weighted belief entropy [91] is employed to measure the uncertainty associated with each body of evidence.

To assign weights, reward and penalty functions derived from IDECA method are used, enabling the expression of the relative significance of each information source. Finally, we apply the classical Dempster's rule to combine the weighted bodies of evidence.

3.4.1 Fuzzy-based similarity measure model

In this section, we introduce a novel approach for measuring conflict by integrating DS evidence theory with a fuzzy inference mechanism. For a thorough and precise evaluation of the conflict level among the pieces of evidence, we suggest employing two metrics:(i) the previously defined improved evidence distance based on Hellinger distance (D_{IH}), and (ii) cosine value (cos). These metrics serve as complementary features for quantifying the degree of conflict, each capturing different aspects of the similarity between Basic Probability Assignments(BPAs). The evidence distance reflects the dissimilarity among the pieces of evidence, and the cosine value provides insights into their consistency. Consequently, we devise a new fuzzy inference mechanism where the evidence distance and the cosine value between BPAs serve as input variables, while the similarity degree between bodies of evidence acts as the output variable, The fuzzy-based similarity measurement model mainly comprises 3 steps; fuzzification, fuzzy rule formulation and inference, and defuzzification. Initially, the conflict measurement factors (D_{IH}, Cos) undergo fuzzification, converting them into fuzzy sets. Subsequently, fuzzy rules are formulated based on empirical and logical inference principles. The fuzzy inference mechanism is then applied to these rules to assess the similarity between the bodies of evidence. Finally, defuzzification is performed to convert the fuzzy output into a precise degree of similarity, as illustrated in Figure 3.3.

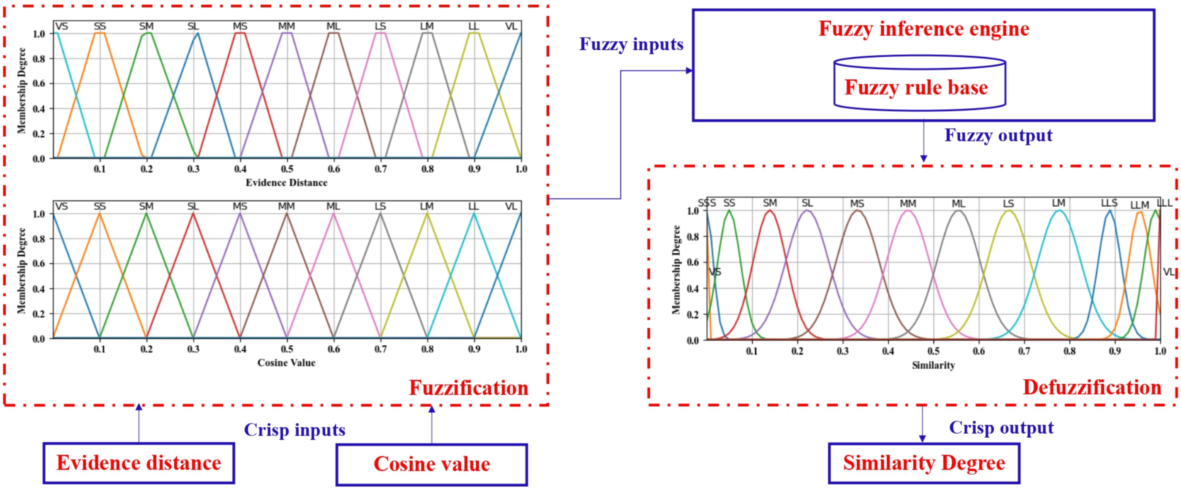


Figure 3.3 – *Fuzzy similarity inference bloc*

3.4.1.1 Fuzzification

The process of fuzzification involves mapping values within the input variables' range to corresponding fuzzy subsets determined by membership functions. Both the evidence distance (D_{IH}) and cosine value (Cos) range from 0 to 1, as does the similarity degree (Sim) between the bodies of evidence. To elucidate the significance of these variables, linguistic descriptors are employed to characterize input variable and output variable traits. The range of the input variables is represented by 11 linguistic terms as very small, very large...etc, while the output variable is represented by 14 linguistic terms to more accurately reflect the mapping distribution between larger and smaller values. Drawing from both data testing and experts' insights, we have determined the relevant parameters for the Trapezoidal, Trigonometric, and Gaussian membership functions corresponding to the evidence distance, cosine value, and similarity degree respectively, as depicted in Table 3.1, Table 3.2 and Table 3.3.

Fuzzy sets	Abbreviations	Trapezoidal MF parameters
Very Small	VS	a = 0, b = 0, c = 0.01, d = 0.09
Small-Small	SS	a = 0.01, b = 0.09, c = 0.11, d = 0.192
Small-Medium	SM	a = 0.11, b = 0.192, c = 0.21, d = 0.3056
Small-Large	SL	a = 0.21, b = 0.3056, c = 0.31, d = 0.39
Medium-Small	MS	a = 0.31, b = 0.39, c = 0.41, d = 0.49
Medium-Medium	MM	a = 0.41, b = 0.49, c = 0.51, d = 0.59
Medium-Large	ML	a = 0.51, b = 0.59, c = 0.61, d = 0.69
Large-Small	LS	a = 0.61, b = 0.69, c = 0.71, d = 0.79
Large-Medium	LM	a = 0.71, b = 0.79, c = 0.81, d = 0.89
Large-Large	LL	a = 0.81, b = 0.89, c = 0.91, d = 1
Very Large	VL	a = 0.9, b = 1, c = 1, d = 1

Table 3.1 – *The Trapezoidal membership function Parameters for evidence distance* d_{IH}

Fuzzy sets	Abbreviations	Trigonometric MF parameters
Very Small	VS	f = 0, m = 0, g = 0.1
Small-Small	SS	f = 0, m = 0.1, g = 0.2
Small-Medium	SM	f = 0.1 , m = 0.2 , g = 0.3
Small-Large	SL	f = 0.2, m = 0.3, g = 0.4
Medium-Small	MS	f = 0.3, m = 0.4, g = 0.5
Medium-Medium	MM	f = 0.4, m = 0.5, g = 0.6
Medium-Large	ML	f = 0.5, m = 0.6, g = 0.7
Large-Small	LS	f = 0.6, m = 0.7, g = 0.8
Large-Medium	LM	f = 0.7, m = 0.8, g = 0.9
Large-Large	LL	f = 0.8, m = 0.9, g = 1
Very Large	VL	f = 0.9, m = 0.9, g = 1

Table 3.2 – *The Trigonometric membership function Parameters for cosine value*

Fuzzy sets	Abbreviations	Gaussian MF parameters
Very Small	VS	c = 0, σ = 0.003
Small-Small-Small	SSS	c = 0, σ = 0.02496
Small-Small	SS	c = 0.005, σ = 0.02496
Small-Medium	SM	c = 0.1326, σ = 0.0366
Small-Large	SL	c = 0.2222, σ = 0.04718
Medium-Small	MS	c = 0.3333, σ = 0.04718
Medium-Medium	MM	c = 0.4444, σ = 0.04718
Medium-Large	ML	c = 0.5555, σ = 0.04718
Large-Small	LS	c = 0.6666, σ = 0.04718
Large-Medium	LM	c = 0.7777, σ = 0.04718
Large-Large-Small	LLS	c = 0.8888, σ = 0.02496
Large-Large-Medium	LLM	c = 0.9555, σ = 0.02496
Large-Large-large	LLL	c = 0.9899, σ = 0.02496
Very Large	VL	c = 1, σ= 0.003

Table 3.3 – *The Gaussian membership function Parameters for similarity degree*

3.4.1.2 Fuzzy rules and inference

As the distance between the pieces of evidence decreases, their similarity increases. Conversely, higher cosine values indicate greater consistency among the pieces of evidence. Leveraging this, fuzzy rules are formulated to articulate the correlation between these metrics in determining the similarity degree between the bodies of evidence. A comprehensive set of 121 Mamdani-type fuzzy rules are devised as conditional statements, listed in Table 3.4. Once the fuzzy rule base is established, Mamdani's maximum operator (logical AND) is applied in fuzzy inference to reason and derive the fuzzy set of the output.

$d_{IH}(m_i, m_j)$ $Cos(m_i, m_j)$	VS	SS	SM	SL	MS	MM	ML	LS	LM	LL	VL
VS	SM	SM	SM	SM	SM	SS	SS	SS	VS	VS	VS
SS	SL	SL	SL	SM	SM	SM	SS	SS	VS	VS	VS
SM	SL	SL	SL	SL	SM	SM	SS	SM	SSS	VS	SSS
SL	MS	MS	MS	MS	SL	SL	SS	SM	SM	SS	SS
MS	MM	MM	MM	MS	MS	SL	SS	SM	SM	SS	SS
MM	ML	ML	ML	MS	MS	MS	SS	SM	SM	SM	SS
ML	ML	ML	ML	MS	MM	MM	SM	SM	SM	SM	SM
LS	LLS	LS	ML	MS	ML	MM	MS	MS	SL	SM	SM
LM	LLM	VL	LS	LS	ML	ML	MM	MS	SL	SL	SM
LL	LLL	VL	LM	LS	ML	ML	MM	MS	SL	SL	SM
VL	VL	VL	LLS	LS	ML	ML	MM	MS	SL	SL	SM

Table 3.4 – *Fuzzy rules*

3.4.1.3 Defuzzification

In this approach, the centroid method is used for defuzzification as defined in Eq. 2.19. It entails identifying the abscissa that represents the center of gravity on the membership function graph associated with the fuzzy set of the output (Similarity Degree).

3.4.2 Conflict measure

According to Eq. 3.11, the enhanced evidence distance is computed between each pair of bodies of evidence $m_i(i = 1, 2, \ldots, N)$ and $m_j(j = 1, 2, \ldots, N)$ to measure the conflict degree. Using Eq. 2.69, the cosine value is calculated for the same pairs of bodies of evidence to quantify the consistency degree. Subsequently, the novel fuzzy-based similarity measure, previously introduced, is applied to determine the overall similarity degree $sim(m_i, m_j)$ between the evidence sources.

3.4.3 Credibility degree

Based on the previously calculated similarity degree $sim(m_i, m_j)$, the support degree $sup(m_i)$ of each evidence can be evaluated using Eq.3.34, and the credibility degree is then obtained using Eq.3.35

3.4.4 Uncertainty measure

To effectively quantify uncertainty, we employ a weighted Deng entropy as outlined in Eq.2.55. Unlike conventional measures, this enhanced entropy considers not only the mass functions but also the proportional scale of propositions or focal elements within the Body of evidence (BOE). By integrating these elements, it offers a comprehensive assessment of uncertain information present in the BOE, thereby addressing uncertainties ignored by standard Deng entropy measures adopted in both AWECA and IDECA approaches.

3.4.5 Weighted BPAs calculation

3.4.5.1 Weights' determination

1. **Initial weights**

   To establish the initial evidence weights, We use penalty and reward functions outlined within IDECA approach, formulated by equations 3.37 and 3.38. These functions are derived independently of a reliability condition, providing a clear and distinct methodology for determining the initial weights.

2. **Final weights**

   The final weights are determined by incorporating both credibility degree and initial weights as defined in Eq.3.39

The final weights are used to modify the original mass functions, the weighted BPAs are then obtained using Eq. 3.40

3.4.6 Evidence combination

Finally, for N body of evidence $m_1, m_2, ..., m_N$, the classical Dempster's combination rule is applied $N-1$ times to fuse the weighted BPAs.

The flowchart of this approach is given in Figure 3.4

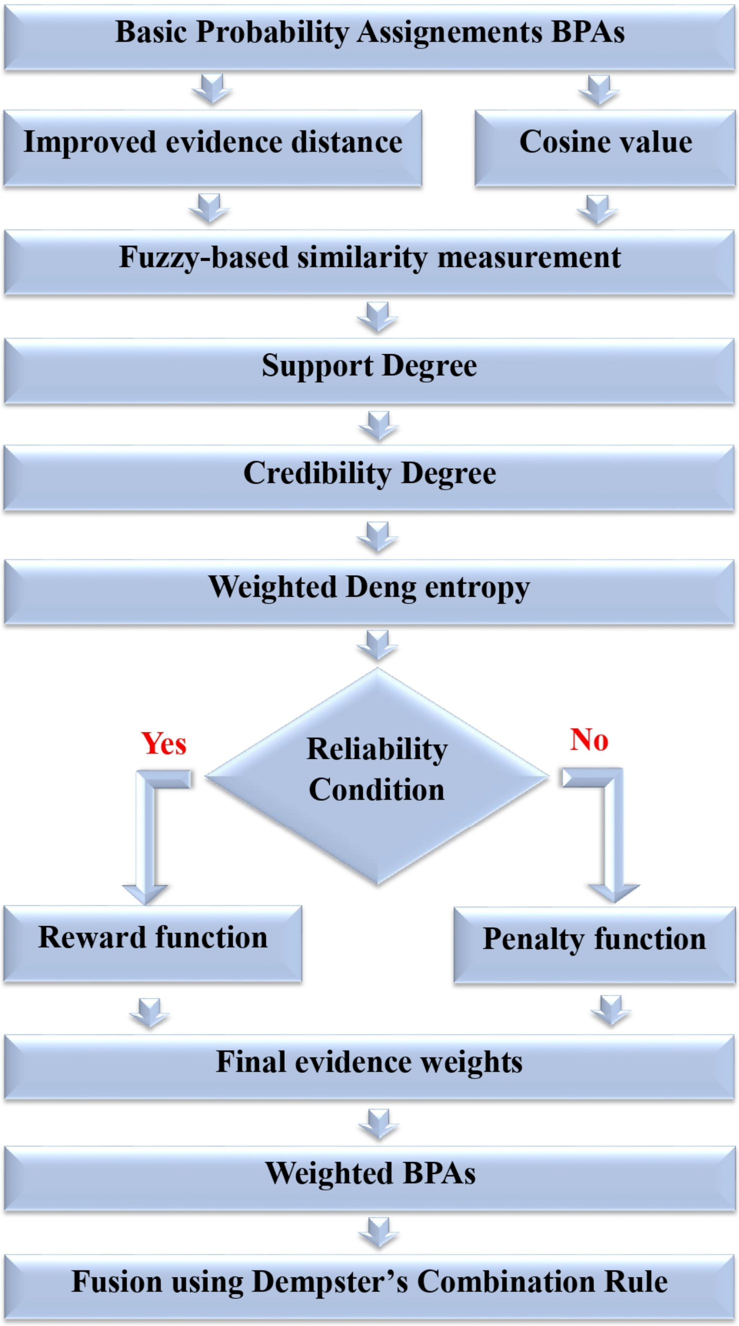


Figure 3.4 – *Flowchart of FSECA method*

3.5 Conclusion

This chapter is dedicated to the development of the improved evidence combination approaches, specifically designed to overcome the conflict problem encountered by Dempster's combination rule. The effectiveness of these proposed approaches will be investigated in the next chapter.

4

Simulation results and discussion

Contents

4.1 Introduction

This part of the present thesis focuses on evaluating the proposed evidence combination approaches built in the previous chapter; common fuse paradox scenarios are used to verify the resilience and effectiveness of the proposed methods, a benchmark numerical example from the literature is used for a comparative study, employing several state of the art methods to highlight the superiority of our approaches. The rationality and validity of our methods across various problem domains, such as fault diagnosis, IoT decision-making, and situational awareness within multi-UAV systems are demonstrated.

4.2 Common conflict paradoxes

We use the common paradoxes outlined in Section 2.7.1 to discuss the rationality and validity of the improved approaches. The BPAs of the various paradox scenarios are depicted in table 4.1.

Paradoxes	Evidence	Propositions				
		A	B	C	D	E
Complete conflict paradox	m_1	1	0	0	/	/
	m_2	0	1	0	/	/
	m_3	0.8	0.1	0.1	/	/
	m_4	0.8	0.1	0.1	/	/
0 Trust paradox	m_1	0.5	0.2	0.3	/	/
	m_2	0.5	0.2	0.3	/	/
	m_3	0	0.9	0.1	/	/
	m_4	0.5	0.2	0.3	/	/
1 Trust paradox	m_1	0.9	0.1	0		
	m_2	0	0.1	0.9	/	/
	m_3	0.1	0.15	0.75	/	/
	m_4	0.1	0.15	0.75	/	/
High conflict paradox	m_1	0.7	0.1	0.1	0	0.1
	m_2	0	0.5	0.2	0.1	0.2
	m_3	0.6	0.1	0.15	0	0.15
	m_4	0.55	0.1	0.1	0.15	0.1
	m_5	0.6	0.1	0.2	0	0.1

Table 4.1 – *BPAs of common conflict paradoxes*

Evidently from table 4.1, the relatively consistent pieces of evidence encompasses:

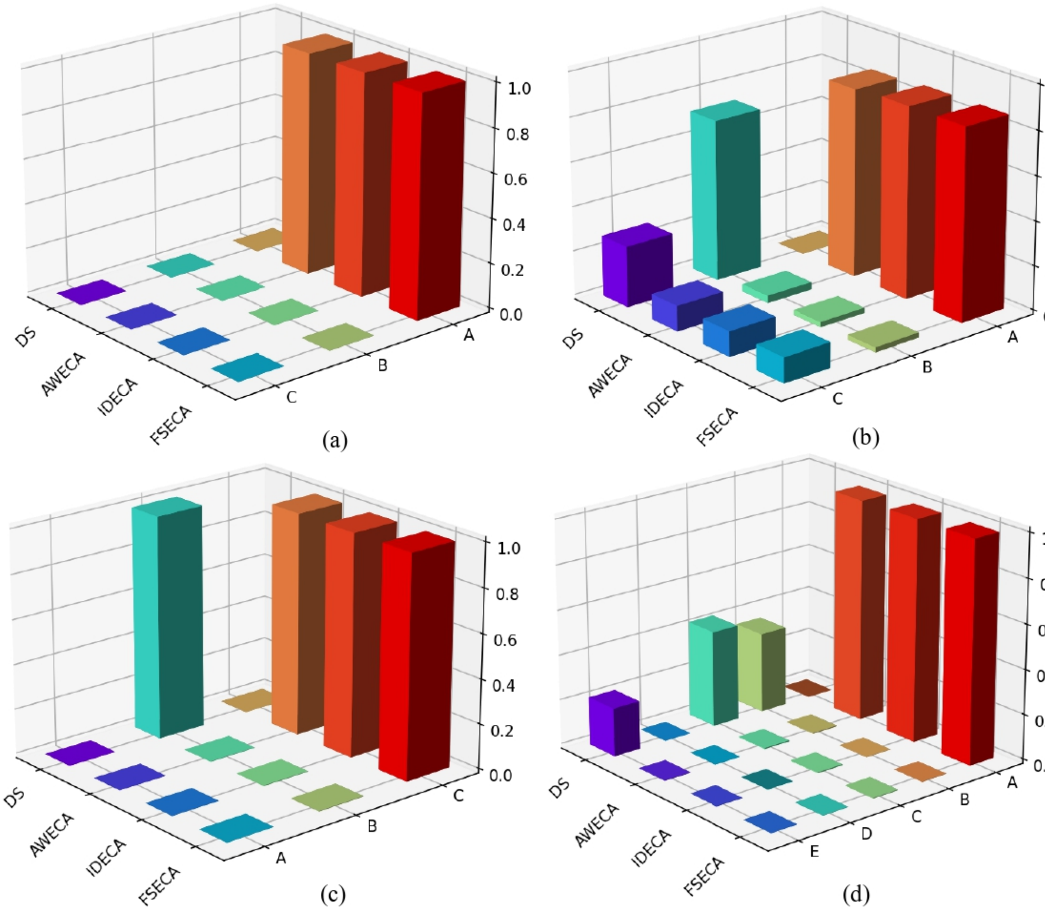
- m_1, m_3, and m_4 in complete conflict paradoxes.
- m_1, m_2, and m_4 in the 0 trust paradox
- m_2, m_3, and $m_4$in the 1 trust paradox
- m_1, m_3, m_4, and m_5 in the high conflict paradox.

Therefore, accurate synthesis results should align with the aforementioned consistent pieces of evidence while avoiding conflicting ones.

Fusion results using DS theory and our improved methods; AWECA, IDECA, FSECA are depicted in table 4.2 and Figure 4.1.

Paradoxes	Methods	Propositions				
		A	B	C	D	E
Complete conflict paradox	Dempster	/	/	/	/	/
	AWECA	0.9998	0.00013	0.000109	/	/
	IDECA	0.9995	0.00025	0.000109	/	/
	FSECA	0.9995	0.00016	0.00016	/	/
0 Trust paradox	Dempster	0	0.727	0.273	/	/
	AWECA	0.8530	0.03284	0.11411	/	/
	IDECA	0.8650	0.022676	0.1123	/	/
	FSECA	0.9997	0.0222	0.000016	/	/
1 Trust paradox	Dempster	0	1	0		
	AWECA	0.000154	0.00106	0.9988	/	/
	IDECA	0.0003247	0.00155	0.9981	/	/
	FSECA	0.00018	0.0013	0.9986	/	/
High conflict paradox	Dempster	0	0.3571	0.4286	0	0.2143
	AWECA	0.9920	0.0029	0.0036	0.0002	0.0014
	IDECA	0.9951	0.00073	0.0034	0	0.00064
	FSECA	0.9961	0.00076	0.00216	0.00076	0.00016

Table 4.2 – *Fusion results of evidence from paradox scenarios using DS theory and our improved approaches*



(a)

(b)

(c)

(d)

Figure 4.1 – *Comparison of the fusion results of the common conflict paradoxes: (a) Complete conflict paradox (b) 0 Trust paradox (c) 1 Trust paradox (d) High conflict paradox*

For the complete conflict paradox, it's evident that Dempster's combination rule cannot be applied to fuse the pieces of evidence. Moreover, it fails to handle all the other three paradoxes effectively, resulting in counterintuitive outcomes by assigning most of the belief to the wrong proposition across all scenarios.

In contrast, our proposed approaches consistently lead to reasonable results. They successfully identify the correct propositions using the maximum BPA with high belief degrees for all paradox situations. These results align with intuitive reasoning and demonstrate the validity and robustness of our proposed approaches — AWECA, IDECA, and FSECA — for all the conflicting situations.

4.3 Benchmark numerical example: Target recognition

In this section, a benchmark numerical example from literature is used to compare our proposed solutions with similar state-of-the-art methods.

In a multisensor-based automatic target recognition system using five different types of sensors, three objects, denoted as A, B, and C, are detected. Let's assume that the frame of discernment, denoted as $\Omega = \{A, B, C\}$, is complete. and A is the current target. Sensor data modeled as BPAs are given in Table 4.3.

	$\{A\}$	$\{B\}$	$\{C\}$	$\{AC\}$
$S_1 : m_1(.)$	0.41	0.29	0.3	0
$S_2 : m_2(.)$	0	0.9	0.1	0
$S_3 : m_3(.)$	0.58	0.07	0	0.35
$S_4 : m_4(.)$	0.55	0.1	0	0.35
$S_5 : m_4(.)$	0.6	0.1	0	0.3

Table 4.3 – *BPAs of the benchmark example (Where AC stands for A, C).*

We apply our developed solutions to fuse the data gathered from the five sensors. The fusion results are subsequently compared to those obtained using various state-of-the-art methods, including Dempster–Shafer method (DS), Murphy [79], Yong [80], Wang et al. [123], Yuan [85], and Yan et al. [95].

Based on the data presented in Table 4.3, it can be observed that S_2 shows a strong conflict with other pieces of evidence, it assigns most of its belief to the wrong target B, while the remaining pieces of evidence mainly support the right target A. This situation may give rise to illogical results after combination using the classical Dempster's rule, ultimately leading to the misidentification of the target.

Table 4.4 presents the fusion results obtained through our proposed approaches as well as the other methods under consideration, for varying numbers of evidence sources. The results of the table are depicted as graphs in Figure 4.2, from which it can be observed that the classical Dempster's combination rule presents a significant limitation in correctly identifying the target when data from all five sensors are combined. The results indicate that it wrongly attributes most of its belief to target C, while the belief assigned to the correct target A remains consistently at 0, regardless of the number of pieces of evidence considered. This discrepancy resulting from the abnormal source S_2, emphasizes the inadequacy of the classical Dempster's combination rule in managing highly conflicting evidence.

Methods	$m_1 - m_2$	$m_1 - m_3$	$m_1 - m_4$	$m_1 - m_5$
D-S	$m(A) = 0$ $m(B) = 0.8969$ $m(C) = 0.1031$	$m(A) = 0$ $m(B) = 0.6575$ $m(C) = 0.3425$	$m(A) = 0$ $m(B) = 0.3321$ $m(C) = 0.6679$	$m(A) = 0$ $m(B) = 0.1422$ $m(C) = 0.8578$
Murphy	$m(A) = 0.0964$ $m(B) = 0.8119$ $m(C) = 0.0917$ $m(AC) = 0$	$m(A) = 0.4619$ $m(B) = 0.4497$ $m(C) = 0.0794$ $m(AC) = 0.0090$	$m(A) = 0.8362$ $m(B) = 0.1147$ $m(C) = 0.0410$ $m(AC) = 0.0081$	$m(A) = 0.9620$ $m(B) = 0.0210$ $m(C) = 0.0138$ $m(AC) = 0.0032$
Yong	$m(A) = 0.1463$ $m(B) = 0.7620$ $m(C) = 0.0917$ $m(AC) = 0$	$m(A) = 0.6021$ $m(B) = 0.2907$ $m(C) = 0.0353$ $m(AC) = 0.0082$	$m(A) = 0.9330$ $m(B) = 0.0225$ $m(C) = 0.0990$ $m(AC) = 0.0092$	$m(A) = 0.9851$ $m(B) = 0.0017$ $m(C) = 0.0096$ $m(AC) = 0.0035$
Yuan	$m(A) = 0.2849$ $m(B) = 0.5306$ $m(C) = 0.1845$ $m(AC) = 0$	$m(A) = 0.8274$ $m(B) = 0.0609$ $m(C) = 0.0986$ $m(AC) = 0.0131$	$m(A) = 0.9596$ $m(B) = 0.0032$ $m(C) = 0.0267$ $m(AC) = 0.0106$	$m(A) = 0.9886$ $m(B) = 0.0002$ $m(C) = 0.0072$ $m(AC) = 0.0039$
Wang et al.	$m(A) = 0.0964$ $m(B) = 0.8119$ $m(C) = 0.0917$ $m(AC) = 0$	$m(A) = 0.6495$ $m(B) = 0.2367$ $m(C) = 0.1065$ $m(AC) = 0.0079$	$m(A) = 0.9577$ $m(B) = 0.0129$ $m(C) = 0.0200$ $m(AC) = 0.0094$	$m(A) = 0.9904$ $m(B) = 0.0009$ $m(C) = 0.0068$ $m(AC) = 0.0019$
Yan et al.	$m(A) = 0.2850$ $m(B) = 0.5310$ $m(C) = 0.1840$ $m(AC) = 0$	$m(A) = 0.08010$ $m(B) = 0.0910$ $m(C) = 0.0950$ $m(AC) = 0.0140$	$m(A) = 0.9460$ $m(B) = 0.0110$ $m(C) = 0.0340$ $m(AC) = 0.0090$	$m(A) = 0.9850$ $m(B) = 0.0010$ $m(C) = 0.0110$ $m(AC) = 0.0030$
AWECA	$m(A) = 0.2678$ $m(B) = 0.5551$ $m(C) = 0.1771$ $m(AC) = 0$	$m(A) = 0.8632$ $m(B) = 0.0449$ $m(C) = 0.0687$ $m(AC) = 0.0231$	$m(A) = 0.9712$ $m(B) = 0.0010$ $m(C) = 0.0142$ $m(AC) = 0.0136$	$m(A) = 0.9904$ $m(B) = 0.0001$ $m(C) = 0.0049$ $m(AC) = 0.0047$
IDECA	$m(A) = 0.2677$ $m(B) = 0.5552$ $m(C) = 0.1770$ $m(AC) = 0$	$m(A) = 0.8800$ $m(B) = 0.0179$ $m(C) = 0.0859$ $m(AC) = 0.0161$	$m(A) = 0.9780$ $m(B) = 0.00009$ $m(C) = 0.0011$ $m(AC) = 0.0208$	$m(A) = 0.9932$ $m(B) = 0.00001$ $m(C) = 0.0002$ $m(AC) = 0.0065$
FSECA	$m(A) = 0.3511$ $m(B) = 0.4368$ $m(C) = 0.2120$ $m(AC) = 0$	$m(A) = 0.8531$ $m(B) = 0.0289$ $m(C) = 0.1075$ $m(AC) = 0.0161$	$m(A) = 0.9783$ $m(B) = 0.0000808$ $m(C) = 0.00026$ $m(AC) = 0.0213$	$m(A) = 0.9933$ $m(B) = 0.00001$ $m(C) = 0.00009$ $m(AC) = 0.0066$

Table 4.4 – *Fusion results by different methods for various numbers of evidence*

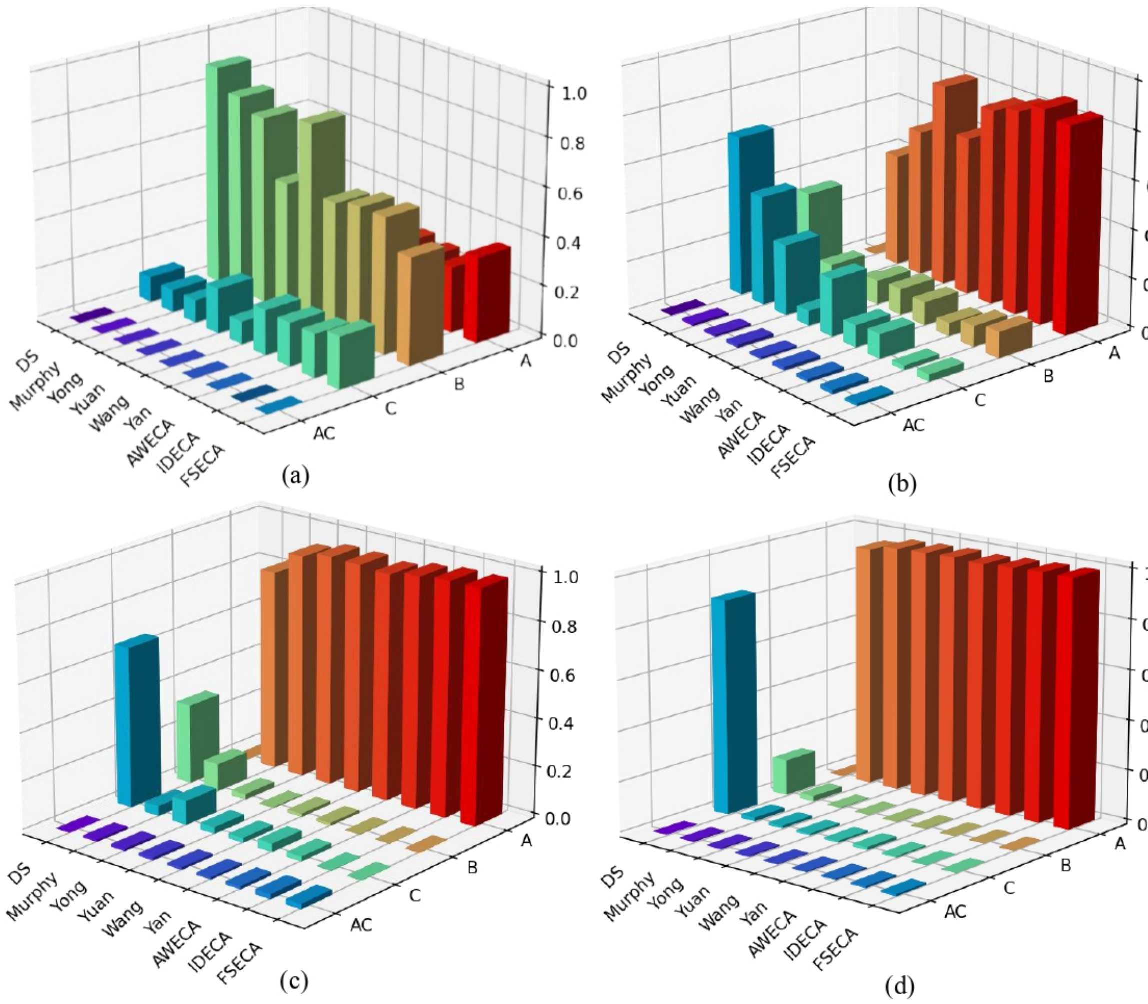


Figure 4.2 – *Fusion results of target recognition application in different methods for various number of evidence: (a) Two pieces of evidence $m_1 - m_2$; (b) Three pieces of evidence $m_1 - m_3$; (c) Four pieces of evidence $m_1 - m_4$; (d) Five pieces of evidence $m_1 - m_5$.*

In the case of the other fusion methods, including our proposed approaches AWECA, IDECA and FSECA, they initially misidentify the target as B when only the two sensors S_1 and S_2 are considered. This misclassification is primarily due to the conflicting evidence S_2, which misguides the fusion process. However, as more sensors are included, along with more reliable pieces of evidence (i.e., S_3, S_4, S_5), all the methods achieve reasonable results and correctly identify the target as A.

Figure 4.3 depicts the evolution of the belief degree assigned to the correct target A by the different methods following each combination involving the five sensors. While all methods ultimately converge to target A as the number of sensors increases, they present distinct rates of convergence and varying levels of belief in the process. As the number of combined sensors increases, the belief degrees of all of our proposed solutions tend to approach 1. When all five pieces of evidence are combined, AWECA reaches a high accuracy rate of 99.04% while IDECA and FSECA achieve even more remarkable accuracy rates of 99.32% and 99.33%, respectively, in correctly identifying the target, outperforming

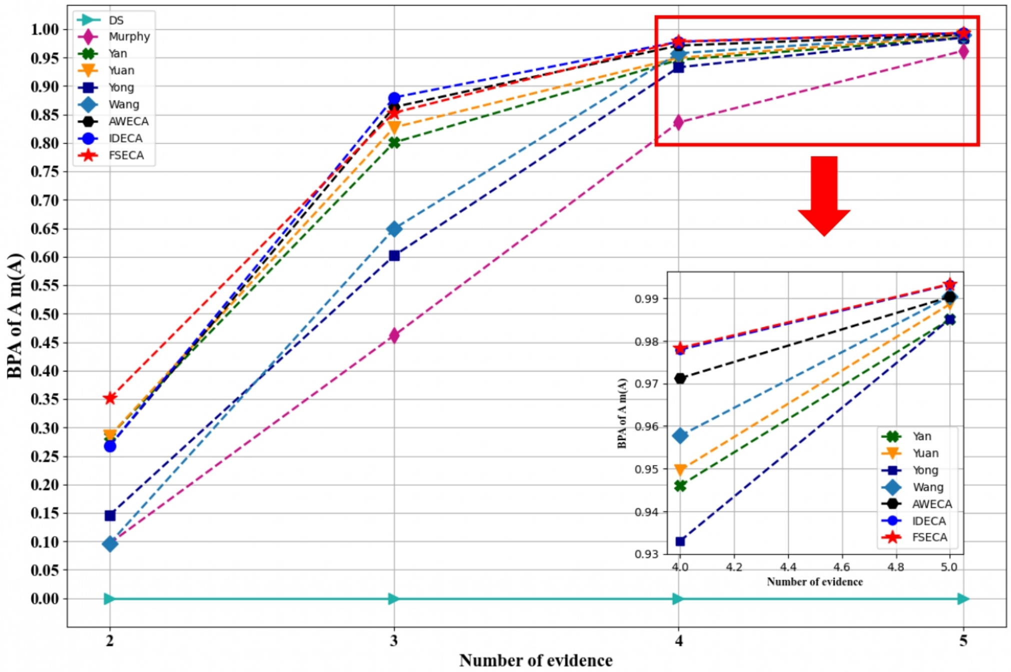


Figure 4.3 – *Comparison of the BPA of target A using different methods*

all competing methods. It is worth highlighting that even a slight increase in accuracy is significant, representing a substantial enhancement in the overall performance. These results validate the effectiveness and superiority of our proposed combination approaches. They adeptly manage conflicts among the pieces of evidence, boasting the best convergence speed and decision accuracy. Our solutions are specifically designed to assess the credibility of each piece of evidence, establish the significance of each sensor in shaping the final fusion result, and assign weights accordingly. This capacity empowers our approaches to mitigate the impact of conflicting pieces of evidence, ultimately leading to enhanced decision accuracy and increased reliability in fusion results.

4.4 Application 1: Fault diagnosis

A case study of a rotating machinery system from [98] is used, where the faults in the system are classified into four types: F_1 = "Imbalance", F_2 = "Shaft crack", F_3 = "Misalignment", and F_4 = "Bearing loose". Therefore, the frame of discernment is built as $\Omega = \{F_1, F_2, F_3, F_4\}$. Five different sensors have been used to monitor the system status and determine the fault type. Data were collected when the fault F_3 was occurring. Fault features were extracted from the data provided by the various sensors. Based on this data, the BPAs (Basic Probability Assignments) of the five sensors were generated. By combining these pieces of evidence, a decision about the system status can be made. Table 4.5 illustrates the calculated BPAs of the different sensors.

	$\{F_1\}$	$\{F_2\}$	$\{F_3\}$	$\{F_4\}$
$S_1 : m_1(.)$	0.1469	0.2057	0.4660	0.1813
$S_2 : m_2(.)$	0.1521	0.1935	0.4631	0.1914
$S_3 : m_3(.)$	0.1278	0.5008	0.2221	0.1493
$S_4 : m_4(.)$	0.1459	0.2396	0.4395	0.1750
$S_5 : m_4(.)$	0.2068	0.1399	0.1755	0.4777

Table 4.5 – *BPAs modeled from the five sensors*

Data provided by S_1, S_2, and S_4 indicate that the fault type is F_3, however S_3 and S_5 provide conflicting information, with S_3 indicating fault type F_2 and S_5 indicating fault type F_4.

We compare the fusion results obtained through our proposed approaches with those derived from various methods, including Dempster Shafer method, Lin [89], Wang [124], and IDCR [98] methods. The fusion results, based on varying number of evidence, are presented in Table 4.6 and illustrated in Figure 4.4.

Methods	$m_1 - m_2$	$m_1 - m_3$	$m_1 - m_4$	$m_1 - m_5$
DS	$m(F_1) = 0.0714$ $m(F_2) = 0.1273$ $m(F_3) = 0.6902$ $m(F_4) = 0.1110$	$m(F_1) = 0.0376$ $m(F_2) = 0.2626$ $m(F_3) = 0.6315$ $m(F_4) = 0.0683$	$m(F_1) = 0.0153$ $m(F_2) = 0.1758$ $m(F_3) = 0.7755$ $m(F_4) = 0.0334$	$m(F_1) = 0.0176$ $m(F_2) = 0.1368$ $m(F_3) = 0.7570$ $m(F_4) = 0.0886$
Lin et al.	$m(F_1) = 0.0715$ $m(F_2) = 0.1274$ $m(F_3) = 0.6903$ $m(F_4) = 0.1111$	$m(F_1) = 0.0315$ $m(F_2) = 0.2675$ $m(F_3) = 0.6431$ $m(F_4) = 0.0579$	$m(F_1) = 0.0125$ $m(F_2) = 0.1692$ $m(F_3) = 0.7906$ $m(F_4) = 0.0276$	$m(F_1) = 0.0109$ $m(F_2) = 0.1258$ $m(F_3) = 0.7874$ $m(F_4) = 0.0759$
Wang et al.	$m(F_1) = 0.0715$ $m(F_2) = 0.1274$ $m(F_3) = 0.6900$ $m(F_4) = 0.1110$	$m(F_1) = 0.0314$ $m(F_2) = 0.2594$ $m(F_3) = 0.6490$ $m(F_4) = 0.0578$	$m(F_1) = 0.0126$ $m(F_2) = 0.1643$ $m(F_3) = 0.8026$ $m(F_4) = 0.0278$	$m(F_1) = 0.0108$ $m(F_2) = 0.1204$ $m(F_3) = 0.7941$ $m(F_4) = 0.0747$
IDCR	$m(F_1) = 0.0715$ $m(F_2) = 0.1274$ $m(F_3) = 0.6901$ $m(F_4) = 0.1111$	$m(F_1) = 0.0315$ $m(F_2) = 0.2540$ $m(F_3) = 0.6565$ $m(F_4) = 0.0585$	$m(F_1) = 0.0124$ $m(F_2) = 0.1571$ $m(F_3) = 0.8029$ $m(F_4) = 0.0275$	$m(F_1) = 0.0103$ $m(F_2) = 0.1148$ $m(F_3) = 0.8011$ $m(F_4) = 0.0692$
AWECA	$m(F_1) = 0.0715$ $m(F_2) = 0.1273$ $m(F_3) = 0.6901$ $m(F_4) = 0.1111$	$m(F_1) = 0.0314$ $m(F_2) = 0.2198$ $m(F_3) = 0.6905$ $m(F_4) = 0.0583$	$m(F_1) = 0.0122$ $m(F_2) = 0.1348$ $m(F_3) = 0.8259$ $m(F_4) = 0.0271$	$m(F_1) = 0.0046$ $m(F_2) = 0.0950$ $m(F_3) = 0.8878$ $m(F_4) = 0.0125$
IDECA	$m(F_1) = 0.0715$ $m(F_2) = 0.1274$ $m(F_3) = 0.6901$ $m(F_4) = 0.1110$	$m(F_1) = 0.0285$ $m(F_2) = 0.0733$ $m(F_3) = 0.8429$ $m(F_4) = 0.0552$	$m(F_1) = 0.0104$ $m(F_2) = 0.0467$ $m(F_3) = 0.9190$ $m(F_4) = 0.0238$	$m(F_1) = 0.0037$ $m(F_2) = 0.0237$ $m(F_3) = 0.9614$ $m(F_4) = 0.0111$
FSECA	$m(F_1) = 0.0715$ $m(F_2) = 0.1274$ $m(F_3) = 0.6901$ $m(F_4) = 0.1111$	$m(F_1) = 0.0283$ $m(F_2) = 0.0684$ $m(F_3) = 0.8484$ $m(F_4) = 0.0549$	$m(F_1) = 0.01033$ $m(F_2) = 0.04505$ $m(F_3) = 0.9210$ $m(F_4) = 0.02365$	$m(F_1) = 0.0035$ $m(F_2) = 0.0222$ $m(F_3) = 0.9642$ $m(F_4) = 0.01008$

Table 4.6 – *Fusion results using different methods.*

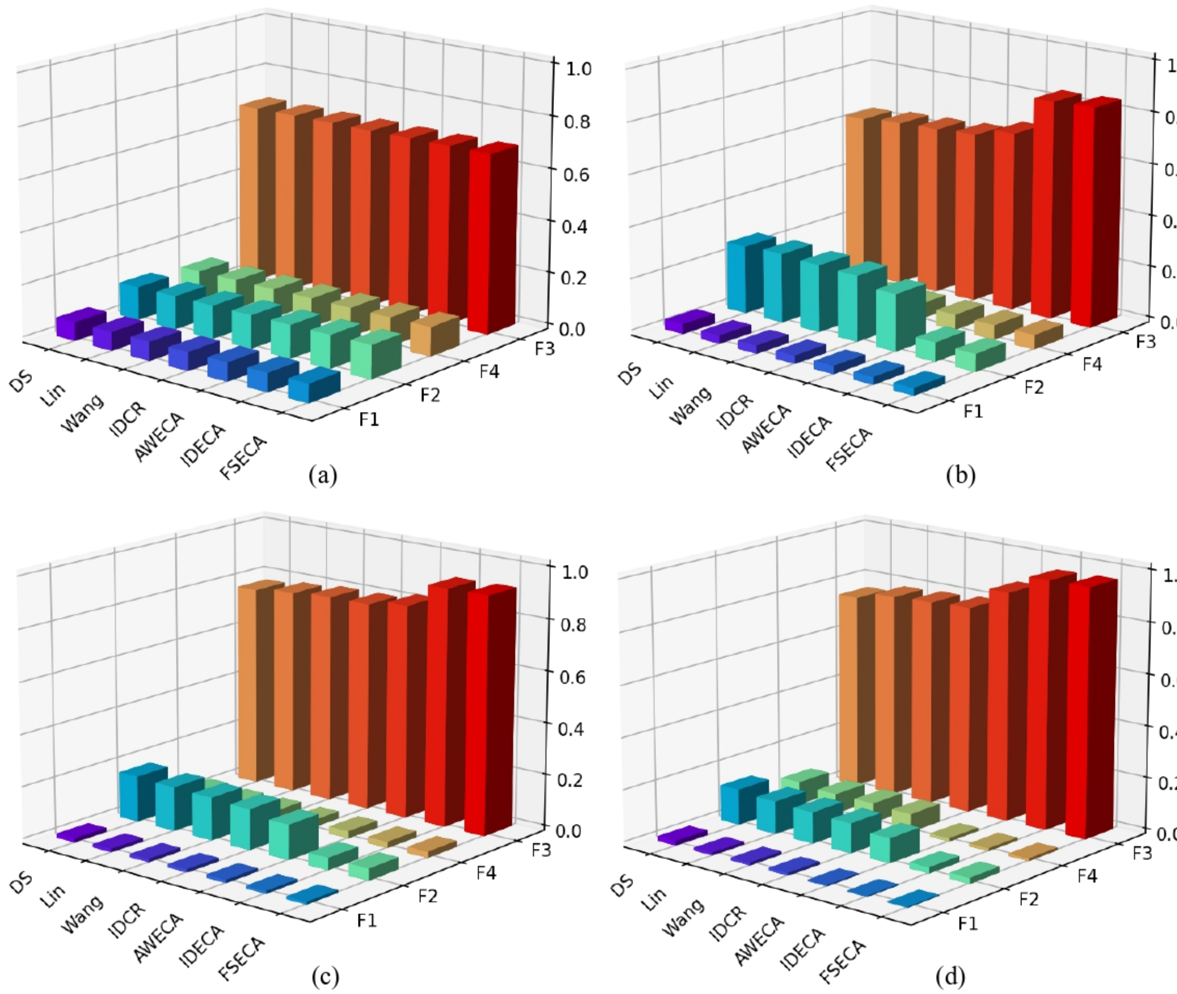


Figure 4.4 – *Fusion results in different methods for various number of evidence: (a) Two pieces of evidence $m_1 - m_2$; (b) Three pieces of evidence $m_1 - m_3$; (c) Four pieces of evidence $m_1 - m_4$; (d) Five pieces of evidence $m_1 - m_5$.*

From Figure 4.4, it is evident that all the methods, including our solutions, successfully diagnose the fault type as F_3 following each combination of five pieces of evidence.

Figure 4.5 depicts the evolution of the belief degree assigned to the right fault type F_3 for the various compared methods, considering varying numbers of sensors. Notably, our proposed approaches consistently outperform all the other methods. When combining three sensors, there is a slight decrease in the belief degree of F_3 for the DS, Lin [89], Wang [124], and IDCR [98] methods. This decline can be attributed to the conflicting data provided by S_3. However, with the inclusion of S_4 in the fusion process, the belief degree of F_3 rises again, reaching 0.7755 for DS, 0.7906 for Lin, 0.8026 for Wang, and 0.8029 for IDCR methods. Nevertheless, this belief degree decreases once again when sensor S_5 is incorporated into the fusion process for all these methods.

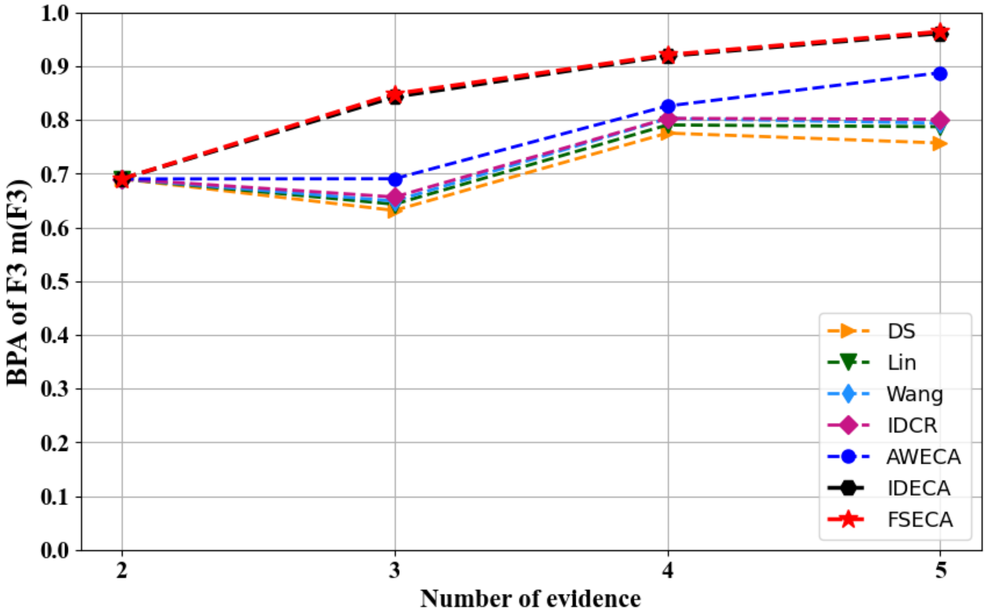


Figure 4.5 – *Comparison of the BPA of the fault F3 for the different methods used*

On the other hand, all of our proposed methods consistently maintain accurate fusion performance. The belief degree assigned to the correct fault type F_3 continues to increase even with the inclusion of the conflicting sources S_3 and S_5. Combining data from all five sensors, AWECA achieves a high accuracy rate of 88.78%, and IDECA achieves 96.14%. Additionally, FSECA reaches the highest belief degree F_3, with an accuracy of 96.42%, surpassing the maximum achieved by any other method which does not exceed 90%.

These results highlight the practical applicability of our proposed solutions, demonstrating their efficiency, validity, and superiority over similar existing approaches.

4.5 Application 2: IoT-based occupancy detection

We examine an IoT decision-making application described in Boulkaboul et al.'s work [61]. The model was assessed through a series of experiments conducted within the context of IoT and smart building projects realized at the CERIST-ALGERIA research center laboratory.

In this scenario, IoT-enabled wireless sensors are deployed to monitor office occupancy and ambient light conditions, to control electrical lighting and optimize energy consumption. Data fusion methods are applied to make informed decisions regarding the activation or deactivation of office lighting. The setup involves the placement of three (03) PIR (Passive Infrared) sensors, denoted as S_1, S_2, and S_3, along with a light sensor labeled as S_4. These sensors are strategically positioned on the office ceiling. Four hypotheses (H_1, H_2, H_3, H_4) based on the office status (occupied or not) and the lighting status (activated or not) are defined as follows:

- H_1: The office is occupied, and the lighting level exceeds 580 lux.
- H_2: The office is unoccupied, but the lighting level exceeds 580 lux.
- H_3: The office is occupied, but the lighting level does not exceed 580 lux.
- H_4: The office is unoccupied, and the lighting level does not exceed 580 lux.

In the baseline scenario, when hypothesis H_1 is confirmed, the system generates evidence indicating the occurrence of hypothesis H_1.

In [61], the influence of the environment on evidence generation was not considered. Consequently, in [125], a belief degree of 10% was assigned to Ω to represent a completely unknown situation. The Basic Probability Assignments (BPAs) derived from the data collected by the four sensors have been calculated using the mean values in each state, they are presented in Table 4.7. The frame of discernment is defined as: $\Omega = \left\{H_1, H_2, H_3, H_4\right\}$

	$\{H_1\}$	$\{H_2\}$	$\{H_3\}$	$\{H_4\}$	$\{\Omega\}$
$S_1 : m_1(.)$	0.729	0.054	0.099	0.018	0.1
$S_2 : m_2(.)$	0.747	0.063	0.081	0.009	0.1
$S_3 : m_3(.)$	0.648	0.153	0.09	0.009	0.1
$S_4 : m_4(.)$	0.621	0.072	0.198	0.009	0.1

Table 4.7 – *BPAs modeled from the four sensors.*

We use our proposed approaches to combine the pieces of evidence and we compare the results obtained with Dempster Shafer method (DS) and four similar existing methods including: Wang[125], Xiao [126], Jiang et al. [127], Wang and Xiao [124] methods. Fusion results are depicted in Table 4.8 and illustrated in Figure 4.6.

Methods	$\{H_1\}$	$\{H_2\}$	$\{H_3\}$	$\{H_4\}$	$\{\Omega\}$
DS	0.9918	0.0027	0.0051	0.0001	0.0003
Wang et al	0.9921	0.0025	0.0050	0.0001	0.0003
Xiao	0.9919	0.0026	0.0051	0.0001	0.0003
Jiang et al	0.9908	0.0030	0.0058	0.0001	0.0003
Wang & Xiao	0.9904	0.0031	0.0061	0.0001	0.0003
AWECA	0.9905	0.003	0.0060	0.0001	0.0003
IDECA	0.9958	0.0012	0.0027	0.0001	0.0002
FSECA	0.9961	0.00108	0.0025	0.0001	0.0002

Table 4.8 – *Fusion results using different methods.*

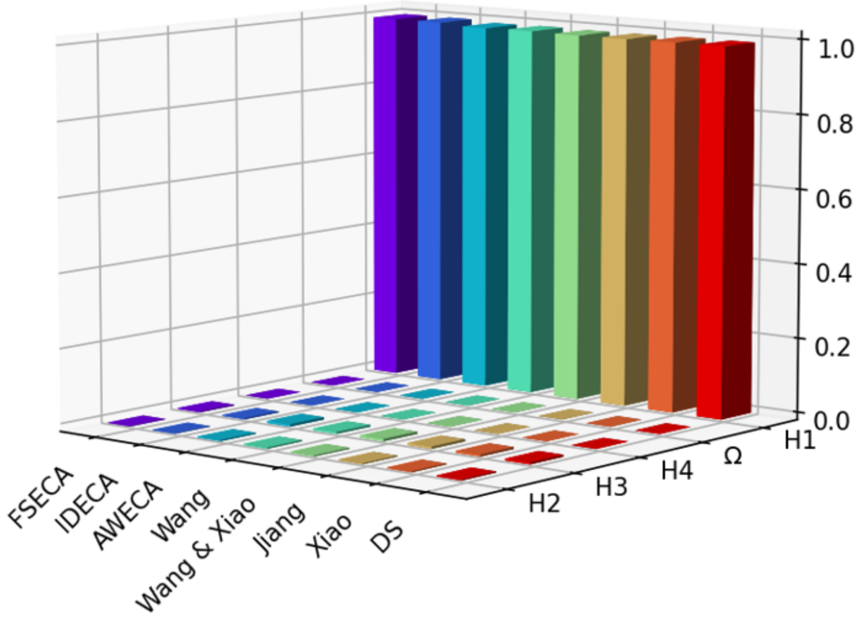


Figure 4.6 – *Comparison of the different methods' fusion results*

As evident from Figure 4.6, all the techniques including Dempster-Shafer method, are capable of identifying the correct hypothesis H_1 , due to the absence of substantial conflicts among the pieces of evidence.

Before combining the pieces of evidence, none of the sensors in Table 4.7 report a belief degree of more than 0.75 for the correct hypothesis H_1 . However, after aplying various combination methods, including our proposed approaches, it is evident that the results converge to yield high belief degrees.

AWECA reaches a high degree of belief of 0.9905 for the right hypothesis H_1 while IDECA and FSECA, deliver significantly better results, providing stronger support for the right hypothesis than the other methods, with a maximum belief degree of 0.9958 and 0.9961, respectively.The superiority of IDECA and FSECA methods stems from the utilization of the proposed evidence distance, which takes into account the relevance and disparities among the pieces of evidence. By detecting even minor conflicts among the pieces of evidence and assigning appropriate weights accordingly, the impact of unreliable pieces of evidence on the final fusion results is minimized. In contrast, the effect of reliable evidence is amplified. This contributes significantly to improved convergence and decision accuracy.

4.6 Application 3: Situational awareness of multi UAV system

4.6.1 Introduction

Situational awareness (SA) of Unmanned Aerial Vehicles (UAVs) refers to their capability to perceive and comprehend their operational environment, including relevant factors such as their position, surroundings, potential threats, and mission objectives. It involves the gathering, processing, and interpretation of data from various sensors and sources to maintain an accurate understanding of the operational context.

By employing multi-sensor data fusion techniques across a network of UAVs, data from various sensors of each UAV can be fused, mitigating uncertainties and enhancing the overall accuracy and reliability of the SA process. This collaborative approach enables UAVs to share and fuse their sensor data in real-time, allowing for a comprehensive understanding of the operational environment. Thus, CSA achieved through the integration of multiple UAVs offers a promising solution to overcome the challenges posed by uncertain environmental conditions, ultimately improving mission effectiveness and safety across a wide range of applications, including surveillance, disaster response, and security operations.

4.6.2 Case study

Data for this case study were provided by [128]. Simulation parameters are presented in Table 4.9 .The system consists of five agents under a specific formation shape, flying forward at a constant speed of 20 m/s along the X-axis and perceiving an obstacle belt, consisting of 50 obstacles. The distribution of both the multi-UAV system and the obstacles is illustrated in Figure 4.7. The data from the UAV system were collected over 20 sampling intervals.

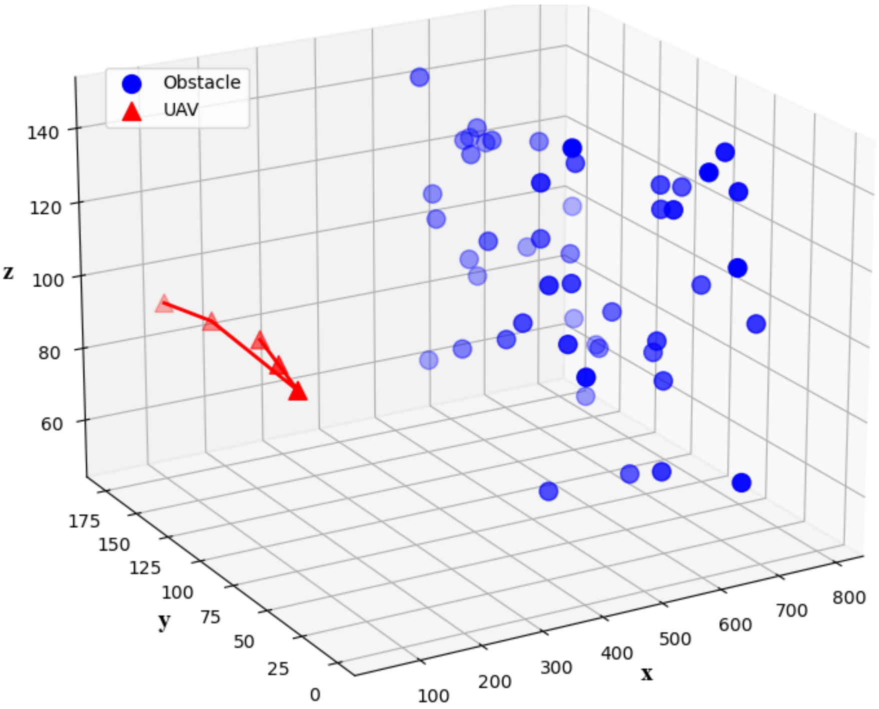


Figure 4.7 – *Distribustion of obstacles and UAVs under a specific formation shape*

Item	Parameter	Item	Parameter
Obstacle space	$300\ m \times 200\ m \times 100\ m$	Δt	$1\ s$
$nOBS$	50	λ	0.6
$nUAV$	5	σ	1
UAV velocity	$20m/s$	α_s	-45° ,+45°
Velocity direction	Positive X axis	L_s	0 - 550m

Table 4.9 – *Similulation parameters*

Note that the formation of the multi-UAV system is fixed, while the obstacles are randomly distributed in a given 3D space. Each UAV is equipped with a depth camera capable of detecting the 3D coordinates of obstacles within a certain range. The information includes the coordinates, relative distance, and visual angle of the obstacle.

Detection capability

The detection range of a depth camera is defined by the maximum visual distance (L_s) and visual angle (α_s) (see figure 4.8). These parameters allow the calculation of the relative distance (L) and visual angle (α) between the UAV and the obstacle.

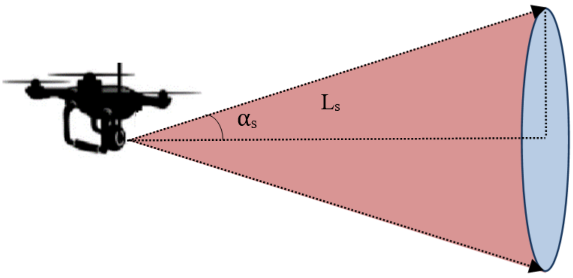


Figure 4.8 – *Depth camera detection range*

Detection accuracy

Obstacles are detectable only within the operational range of sensors and the uncertainty in detection occurs, particularly when an obstacle deviates from the center of sight or is distant from the UAV's depth camera.

This has been modeled by introducing the step function, $\sigma(L,\alpha)$. The function assigns a value of 1 when the obstacle is within the detection range of the sensors. Otherwise, it assigns a value of 0, indicating that the obstacle is outside the detection range and cannot be detected.

$$\sigma(L,\alpha) = \begin{cases} 1,\ if\ -L_s \leqslant L \leqslant L_s\ and\ -\alpha_s \leqslant \alpha \leqslant \alpha_s \\ 0,\ otherwise \end{cases} \tag{4.1}$$

Where the relative distance L and visual angle α between UAV and obstacle can be calculated using the position coordinates of the UAV and the obstacle.

Detection uncertainty

Detection uncertainty arises in depth cameras due to inherent characteristics, manufacturing errors, and assembly imperfections, leading to inaccuracies in obstacle detection.

The detection accuracy diminishes as the visual distance (L) increases and is also influenced by angular offset, where accuracy decreases with higher visual angles α. It has been demonstrated in[128] that the detection accuracy of common depth cameras, exponentially decreases as visual distance increases. Moreover, the detection accuracy exhibits a specific distribution concerning the visual angle, approximating a Gaussian distribution. Thus, the relationship between detection accuracy and visual distance follows a Gaussian distribution.

To describe these relationships mathematically, the study adopts an exponential function $P(L)$ for the relationship between detection accuracy and visual distance, with the probability density function defined as $f(L)$. Similarly, a Gaussian distribution function $P(\alpha)$ is used to represent the relationship between detection accuracy and visual angle, with the probability density function defined as $f(\alpha)$. These mathematical expressions are as follows:

$$f(L) = -\lambda e^{-\lambda L} \tag{4.2}$$

$$f(\alpha) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{\alpha^2}{2\sigma^2}} \tag{4.3}$$

The static detection accuracy is then defined as:

$$\begin{aligned} P(L,\alpha) &= \sqrt{2\pi}\frac{\sigma}{\lambda}. \frac{e^{-\dfrac{\alpha^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}.\lambda e^{-\lambda L}.\sigma(L,\alpha) \\ &= e^{-\dfrac{\alpha^2}{2\sigma^2} - \lambda L}.\sigma(L,\alpha) \end{aligned} \tag{4.4}$$

Where λ and σ represent the uncertainty coefficient of visual distance and angle, respectively.

By considering the time dimension, the BPA m in evidence theory is formulated as follows:

$$\begin{aligned} m(L_t,\alpha_t) &= \sqrt{2\pi}\frac{\sigma}{-}. \frac{e^{-\dfrac{\alpha_t^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}.\lambda e^{-\lambda L_t}.\sigma(L_t,\alpha_t) \\ &= e^{-\dfrac{\alpha_t^2}{2\sigma^2} - \lambda L_t}.\sigma(L_t,\alpha_t) \end{aligned} \tag{4.5}$$

4.6.3 Fusion process

Scenario: We consider the detection of the obstacle Φ_1.

4.6.3.1 modeling

The frame of discernment is designed as follows:

$$\Omega = \{Detected,\ Not\ detected\} \tag{4.6}$$

Mass functions' calculation

Based on the data provided by the 5 UAVs, equation 4.5 is used to calculate the Basic Probability assignments, corresponding to each obstacle, during 20 sampling times.

4.6.3.2 Combination

The mass functions of the 5 UAVs for each obstacle, during 20 sampling times are fused using our improved approaches; AWECA, IDECA, FSECA. Fusion results of detection accuracy for obstacle ϕ_1 are depicted in Figure 4.9.

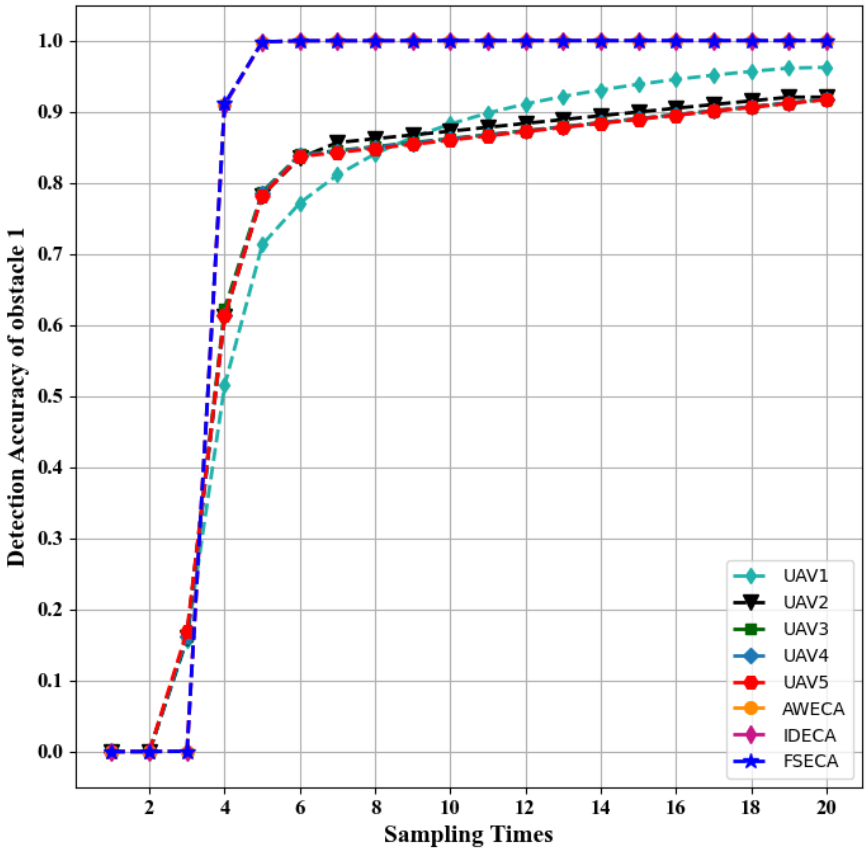


Figure 4.9 – *Comparaison of the detection accuracy between the multi UAV system and the five single UAVs for the obstacle* ϕ_1

According to Figure 4.9, UAV1 starts detecting the obstacle at the third sampling time, initially achieving an accuracy of 15.69%. This accuracy gradually improves, reaching 96.24% by the 20th sampling time. Similarly, UAV2 detects the obstacle from the third sampling time, with an initial accuracy of 15.96%, increasing to 92.04% by the 20th sampling time. UAV3 also detects the obstacle from the third sampling time, starting with an accuracy of 16.23% and improving to 91.77% by the 20th sampling time. UAV4 detects the obstacle from the third sampling time as well, initially achieving an accuracy of 16.28% and improving to 91.79% by the 20th sampling time. Finally, UAV5 detects the obstacle starting from the third sampling time, with an initial accuracy of 16.92% , which improves to 91.69% by the 20th sampling time.

Following the fusion of mass functions from the 5 UAVs over 20 sampling times using our developed approaches, the multi-UAV system starts detecting the obstacle with high accuracy, surpassing 90% , from the fourth sampling time for all our proposed methods. Subsequently, it achieves a maximum detection accuracy of 99.99% after the fifth sampling time, maintaining this level of accuracy until the 20th sampling time.

Notably, the detection accuracy achieved by fusing data from various UAVs using our methods surpasses that achieved by any single sensor alone. Additionally, the graph shows that the performance curves of our proposed methods—AWECA, IDECA, and FSECA—are nearly identical, indicating that they perform almost the same. These results validate the effectiveness of our proposed approaches—AWECA, IDECA, and FSECA— in enhancing UAV situational awareness

4.7 Conclusion

In this chapter, the simulations' results have been presented and analyzed. These results have shown the validity and effectiveness of the built combination strategies in handling conflicts and uncertainties among the evidence sources. Comparisons with various methods from the existing literature have demonstrated the efficiency and superiority of our proposed improved combination approaches in terms of conflict management, convergence speed, reliability of fusion results, and decision accuracy.

General Conclusion

In IoT-based smart environments, vast amounts of data are generated every second. Due to multiple factors, these data are susceptible to various imperfections, such as uncertainty, conflicts, or inaccuracies, potentially leading to erroneous decisions. Multisensor data fusion has emerged as a potent solution for managing data from diverse sources and facilitating effective decision-making.

The Dempster–Shafer (D–S) theory stands out as a robust and flexible mathematical framework for modeling and merging uncertain, imprecise, and incomplete data. Widely applied in multisensor data fusion scenarios, it plays a pivotal role in applications such as decision-making, fault diagnosis, and pattern recognition.

However, the challenge arises when dealing with contradictory data. Combining conflicting sources can be intricate, and the results may be deemed unreasonable. This challenge underscores the need for careful consideration and specialized techniques when handling highly conflicting data sources.

The problem addressed in this thesis is primarily related to the challenges associated with information quality in the context of the Internet of Things (IoT). The objective is to propose alternative operators for combining data gathered from diverse sources within the Dempster-Shafer framework, to address the limitations of the classical Dempster's combination rule and effectively fuse highly conflicting evidence without yielding counter-intuitive results.

To this end, enhanced evidence combination approaches have been developed, they are based on preprocessing the mass functions before the combination. Weights representing the degree of confidence given to data sources are determined using various factors. The methods are specially designed for uncertainty measure and evidence conflict management, enabling the fusion system to reach effective decision-making.

(i) The advanced weighted evidence combination approach (AWECA) primarily relies on three key tools: evidence distance, evidence angle, and belief entropy. Evidence distance measures the dissimilarity between the bodies of evidence, while evidence angle characterizes their consistency. Both metrics quantify the degree of conflict among the sources. On the other hand, belief entropy for measuring the level of uncertainty within each body of evidence. These metrics are then used in the determination of relevant weighting factors. Higher weights are assigned to sources well-supported by others, while lower weights are assigned to less-supported sources, mitigating their conflicting impact on the final fusion results.

(ii) The improved evidence distance-based approach (IDECA): a newly defined improved evidence distance based on Hellinger distance is firstto effectively quantifies the degree of conflict be- tween the bodies of evidence. It takes into account the interdependencies between these pieces of evidence through Jaccard matrix, satisfying key metric properties (non-negativity, symmetry, positive definiteness, trigonometric inequality) while providing a better measure of conflict. Then, a novel evidence fusion strategy is built upon the improved evidence distance to address the conflict degree between the bodies of evidence and it utilizes Deng entropy to quantify the uncertainty associated with each body of evidence. To assign weights, a reliability condition is established, wherein reward and penalty functions are devised. Reliable pieces of evidence receive more significant weights, amplifying their influence on the final fusion result, while less reliable ones are assigned lower weights. Finally, the classical Dempster's rule is employed to combine the modified bodies of evidence.

(iii) Fuzzy-based similarity measure combination method (FSECA), The main concept of this method is to integrate the fuzzy inference mechanism into the similarity measure model to effectively quantify the degree of conflict among the pieces of evidence, using the introduced enhanced distance and cosine value. Expanding on this, a weighted belief entropy is employed to measure the uncertainty associated with each body of evidence. To assign weights, reward and penalty functions derived from IDECA method are used, enabling the expression of the relative significance of each information source. Finally, the classical Dempster's rule to combine the weighted bodies of evidence.

The validity and effectiveness of the proposed approaches have been demonstrated through multiple simulations. First, common fuse paradox scenarios were used to verify the robustness and efficiency of the proposed methods in eliminating the problem of counterintuitive results encountered by Dempster's combination rule when dealing with highly conflicting sources. Our approaches proved to be powerful in addressing these issues. Second, a benchmark numerical example from the literature was used for a comparative study with several state-of-the-art methods. The fusion results showed that our solutions outperformed all other methods in terms of conflict management, convergence speed, and fusion result accuracy. The applicability and validity of our methods across different problem domains, including fault diagnosis, IoT-occupancy detection, and situational awareness within multi-UAV systems, were demonstrated.

Therefore, the obtained results and conclusions that have been derived from this work can confirm that the study may serve as a robust, effective, and accurate solution for multisensor data fusion in IoT environments. All of this, in addition to the rich bibliography that it provides, and that may help as a departure point for further research in the same field.

Future directions

The study of effectively managing the uncertainties and addressing the conflict remains open, with ample room for improvement. In this regard, the following perspectives outline potential axes for future researches:

- **Adaptation to open-world Assumptions**: Our proposed methods are constrained to closed-world scenarios. Investigate adapting these approaches to produce reliable results within an open-world assumption, broadening their applicability.
- **Incorporating Additional Factors**: Enhancing proposed methods by integrating relevant factors such as data timeliness.
- **Unified Methodology for Mass Functions**: Developement of a unified methodology for determining mass functions, providing a standardized framework that promotes consistency and comparability across different applications and domains.
- **Integration of Deep learning techniques with Dempster-Shafer Theory** :Explore the synergy between machine learning techniques and Dempster-Shafer theory, both for quantifying uncertainty and conflict and for facilitating autonomous decision-making.
- **Use of larger Datasets**: Applying the proposed approaches to large datasets enhance their robustness and generalizability, enabling more comprehensive analysis and validation.

Contributions

The following papers are the fruit of our work during the dissertation:

Journal papers

- Hamda, Nour El Imane, Allel Hadjali, and Mohand Lagha. 2023. "Multisensor Data Fusion in IoT Environments in Dempster–Shafer Theory Setting: An Improved Evidence Distance-Based Approach" Sensors 23, no. 11: 5141. https://doi.org/10.3390/s23115141

Conference papers

- Hamda, Nour El Imane, Allel Hadjali, and Mohand Lagha. "An Advanced Weighted Evidence Combination Method for Multisensor Data Fusion in IoT," 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 2022, pp. 810-815, doi: 10.1109/DASA54658.2022.9765125.
- Hamda, Nour El Imane, Mohand Lagha and Allel Hadjali. 2022. " Mathematical Methods for Data Fusion in IoT: A Survey". In: Kacprzyk, J., Balas, V.E., Ezziyyani, M. (eds) Advanced Intelligent Systems for Sustainable Development (AI2SD'2020). AI2SD 2020. Advances in Intelligent Systems and Computing, vol 1418. Springer, Cham. https://doi.org/10.1007/978-3-030-90639 9-88

Bibliography

[1] James Llinas and David L. Hall. Introduction to multi-sensor data fusion. In *IEEE Int. Symp. Circuits Syst*, volume 6, pages 537–540, 1998.

[2] Arthur P. Dempster. A generalization of bayesian inference. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*, 1968.

[3] Glenn Shafer. A mathematical theory of evidence. *NJ: Princeton University Press*, 1976.

[4] Chao Fu and Shanlin Yang. Conjunctive combination of belief functions from dependent sources using positive and negative weight functions. *Expert Systems with Applications*, 41(4, Part 2):1964–1972, 2014.

[5] Liguo Fei, Jun Xia, Yuqiang Feng, and Luning Liu. An electre-based multiple criteria decision making method for supplier selection using dempster-shafer theory. *IEEE Access*, 7:84701–84716, 2019.

[6] Wenjun Ma, Weiru Liu, Xudong Luo, Kevin McAreavey, Yuncheng Jiang, and Jianbing Ma. A dempster-shafer theory and uninorm-based framework of reasoning and multiattribute decision-making for surveillance system. *International Journal of Intelligent Systems*, 34:3077 – 3104, 2019.

[7] Yukun Dong, Jiantao Zhang, Z. Li, Yong Hu, and Yong Deng. Combination of evidential sensor reports with distance function and belief entropy in fault diagnosis. *Int. J. Comput. Commun. Control*, 14:329–343, 2019.

[8] Yan Shu-fa, Ma Biao, Zheng Chang-song, and Chen Man. Weighted evidential fusion method for fault diagnosis of mechanical transmission based on oil analysis data. *International Journal of Automotive Technology*, 20(5):989–996, 2019.

[9] Xianfeng Fan and Ming J. Zuo. Fault diagnosis of machines based on d–s evidence theory. part 1: D–s evidence theory and its improvement. *Pattern Recognition Letters*, 27(5):366–376, 2006.

[10] Ganggang Dong and Gangyao Kuang. Target recognition via information aggregation through dempster–shafer's evidence theory. *IEEE Geoscience and Remote Sensing Letters*, 12(6):1247–1251, 2015.

[11] Dominique Gruyer, Sébastien Demmel, Valentin Magnier, and Rachid Belaroussi. Multi-hypotheses tracking using the dempster–shafer theory, application to ambiguous road context. *Information Fusion*, 29:40–56, 2016.

[12] Tang Yongchuan, Wu Dongdong, and Liu Zijing. A new approach for generation of generalized basic probability assignment in the evidence theory pattern analysis and applications. *Information Fusion*, 24:1007–1023, 2021.

[13] Ovidiu Vermesan, Peter Friess, Patrick Guillemin, Sergio Gusmeroli, Harald Sundmaeker, Alessandro Bassi, Ignacio Soler Jubert, Margaretha Mazura, Mark Harrison, Markus Eisenhauer, and Pat Doody. Internet of things strategic research roadmap. In *Internet of Things Strategic Research Roadmap*, pages 9–52, 2009.

[14] 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. The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015).

[15] Zainab Hassan Ali, Hesham Arafat Ali, and Mahmoud Mohammed Badawy. Internet of things (iot): Definitions, challenges and recent research directions. *International Journal of Computer Applications*, 128:37–47, 2015.

[16] Kevin Ashton. That “internet of things” thing: In the real world things matter more than ideas. *RFID Journal*, 1999.

[17] Pierre-Jean ; BENGHOZI, Sylvain ; BUREAU, and Françoise MASSIT-FOLLÉA. *The Internet of Things What Challenges for Europe?* Paris: Editions of the House of Human Sciences, 2009.

[18] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things: A survey. *Computer Networks*, 54(15):2787–2805, 2010.

[19] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645–1660, 2013.

[20] International telecommunication union ITU-T Y.2060. Next generation networks frameworks and functional architecture models: Overview of the internet of things. 2012.

[21] Rosilah Hassan, Faizan Qamar, Mohammad Kamrul Hasan, Azana Hafizah Mohd Aman, and Amjed Sid Ahmed. Internet of things and its applications: A comprehensive survey. *Symmetry*, 12(10):1674, Oct 2020.

[22] Marco Lombardi, Francesco Pascale, and Domenico Santaniello. Internet of things: A general overview between architectures, protocols and applications. *Information*, 12(2):87, Feb 2021.

[23] Rwan Mahmoud, Tasneem Yousuf, Fadi A. Aloul, and Imran Ahmed Zualkernan. Internet of things (iot) security: Current status, challenges and prospective measures. *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 336–341, 2015.

[24] Mahesh Kavre, Aditya Gadekar, and Yash Gadhade. Internet of things (iot): A survey. In *2019 IEEE Pune Section International Conference (PuneCon)*, pages 1–6, 2019.

[25] Apostolos Gerodimos, Leandros Maglaras, Mohamed Amine Ferrag, Nick Ayres, and Ioanna Kantzavelou. Iot: Communication protocols and security threats. *Internet of Things and Cyber-Physical Systems*, 3:1–13, 2023.

[26] Miguel A. Becerra, Catalina Tobón, Andrés Eduardo Castro-Ospina, and Diego H. Peluffo-Ordóñez. Information quality assessment for data fusion systems. *Data*, 6(6):60, Jun 2021.

[27] Prashanth Southekal. *Data quality*. John Wiley Sons, 2023.

[28] George M. Marakas James A. O'Brien. *Introduction to Information Systems*. McGraw-Hill/Irwin: New York, NY, USA, 2005.

[29] Richard Y. Wang and Diane M. Strong. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst.*, 12:5–33, 1996.

[30] James R. Evans and William M. Lindsay. The management and control of quality. 1989.

[31] Fatimah Sidi, Payam Hassany Shariat Panahy, Lilly Suriani Affendey, Marzanah A. Jabar, Hamidah Ibrahim, and Aida Mustapha. Data quality: A survey of data quality dimensions. In *2012 International Conference on Information Retrieval Knowledge Management*, pages 300–304, 2012.

[32] Laney Douglas. 3D data management: Controlling data volume, velocity, and variety. Technical report, META Group, 2001.

[33] Hamidur Rahman, Shahina Begum, and Mobyen Uddin Ahmed. Ins and outs of big data: A review. In Mobyen Uddin Ahmed, Shahina Begum, and Wasim Raad, editors, *Internet of Things Technologies for HealthCare*, pages 44–51. Springer International Publishing, 2016.

[34] Rajalakshmi Krishnamurthi, Adarsh Kumar, Dhanalekshmi Gopinathan, Anand Nayyar, and Basit Qureshi. An overview of iot sensor data processing, fusion, and analysis techniques. *Sensors*, 20(21):6076, Oct 2020.

[35] Hugh F. Durrant-Whyte and Thomas C. Henderson. Multisensor data fusion. In *Springer Handbook of Robotics, 2nd Ed.*, 2008.

[36] Irwin R. Goodman and Hung T. Nguyen. Uncertainty models for knowledge-based systems. 1985.

[37] Muwei Jian, Junyu Dong, and Yang Zhang. Image fusion based on wavelet transform. In *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007)*, volume 1, pages 713–718, 2007.

[38] K. S. Nagla, Moin Uddin, and Dilbag Singh. Multisensor data fusion and integration for mobile robots: A review. In *IEEE International Conference on Robotics and Automation*, 2014.

[39] Stanislav Vechet and Jiri Krejsa. *Sensors Data Fusion via Bayesian Network*, pages 221–226. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[40] Iván Cabria and Iker Gondra. Mri segmentation fusion for brain tumor detection. *Information Fusion*, 36:1–9, 2017.

[41] L. Gautier, Abdelmalik Taleb-Ahmed, M. C. Rombaut, Jack-Gérard Postaire, and Herve Leclet. Belief function in low level data fusion: application in mri images of vertebra. *Proceedings of the Third International Conference on Information Fusion*, 1, 2000.

[42] Salah Sukkarieh, Eric Nettleton, Jong-Hyuk Kim, Matthew Ridley, Ali Goktogan, and Hugh Durrant-Whyte. The anser project: Data fusion across multiple uninhabited air vehicles. *The International Journal of Robotics Research*, 22(7-8):505–539, 2003.

[43] F. Schettini, Gianpietro Di Rito, Roberto Galatolo, and Eugenio Denti. Sensor fusion approach for aircraft state estimation using inertial and air-data systems. *2016 IEEE Metrology for Aerospace (MetroAeroSpace)*, pages 624–629, 2016.

[44] Kehua Guo, Yayuan Tang, and Peiyun Zhang. Csf: Crowdsourcing semantic fusion for heterogeneous media big data in the internet of things. *Information Fusion*, 37:77–85, 2017.

[45] Min Huang, Zhen Liu, and Yang Tao. Mechanical fault diagnosis and prediction in iot based on multi-source sensing data fusion. *Simulation Modelling Practice and Theory*, 102:101981, 2020.

[46] Lars Zimmermann, Robert Weigel, and Georg Fischer. Fusion of nonintrusive environmental sensors for occupancy detection in smart homes. *IEEE Internet of Things Journal*, 5:2343–2352, 2018.

[47] Lan Zhang, Henry Leung, and Keith C.C. Chan. Information fusion based smart home control system and its application. *IEEE Transactions on Consumer Electronics*, 54(3):1157–1165, 2008.

[48] Rustem Dautov, Salvatore Distefano, and Rajkumar Buyya. Hierarchical data fusion for smart healthcare. *Journal of Big Data*, 6:1–23, 2019.

[49] Ekta Maini, Silpa Ajith Kumar, and D. S. Sagar. Role of data fusion in intelligent transportation system: A survey. 2017.

[50] Andrei B. Bezerra Torres, Atslands Rego da Rocha, Ticiana L. Coelho da Silva, José Neuman de Souza, and Rubens S. Gondim. Multilevel data fusion for the internet of things in smart agriculture. *Comput. Electron. Agric.*, 171:105309, 2020.

[51] Nour El Imane Hamda, Mohand Lagha, and Allel Hadjali. Mathematical methods for data fusion in iot: A survey. In Janusz Kacprzyk, Valentina E. Balas, and Mostafa Ezziyyani, editors, *Advanced Intelligent Systems for Sustainable Development (AI2SD'2020)*, pages 1084–1101, Cham, 2022. Springer International Publishing.

[52] Federico Castanedo. A review of data fusion techniques. *The Scientific World Journal*, 2013, 2013.

[53] Jitendra R. Raol. *Multi-Sensor Data Fusion with MATLAB*. CRC Press, 2009.

[54] Harvey B. Mitchell. Data fusion. In *Data Fusion: Concepts and Ideas*.

[55] Frank White. Data fusion lexicon. In *The Data Fusion Subpanel of the Joint Directors of Laboratories, Technical Panel for C3, Naval Ocean Systems Center, San Diego, Calif, USA*, 1991.

[56] Erik Blasch, James Llinas, Dale Lambert, Pierre Valin, Subrata Das, Chee Chong, Mitch Kokar, and Elisa Shahbazian. High level information fusion developments, issues, and grand challenges: Fusion 2010 panel discussion. In *2010 13th International Conference on Information Fusion*, pages 1–8, 2010.

[57] Shweta Sinha and Priyanka Vashisht. *Explainable Data Fusion on Edge: Challenges and Opportunities*, pages 117–138. Springer International Publishing, Cham, 2023.

[58] Ahmed Abdelgawad and Magdy Bayoumi. *Data Fusion in WSN*, pages 17–35. Springer US, Boston, MA, 2012.

[59] R.C. Luo and M.G. Kay. Multisensor integration and fusion in intelligent systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5):901–931, 1989.

[60] Belur V. Dasarathy. Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proceedings of the IEEE*, 85(1):24–38, 1997.

[61] Sahar Boulkaboul and Djamel Djenouri. Dfiot: Data fusion for internet of things. *Journal of Network and Systems Management*, 28:1136–1160, 2020.

[62] Furqan Alam, Rashid Mehmood, Iyad Katib, Nasser N. Albogami, and Aiiad Albeshri. Data fusion and iot for smart ubiquitous environments: A survey. *IEEE Access*, 5:9533–9554, 2017.

[63] Xiaoming Chen and Xin Li. Virtual temperature measurement for smart buildings via bayesian model fusion. In *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 950–953, 2016.

[64] Waleed A. Abdulhafiz and Alaa Khamis. Bayesian approach to multisensor data fusion with pre- and post-filtering. In *2013 10th IEEE International conference on networking, sensing and control(ICNSC)*, pages 373–378, 2013.

[65] Naiwei Cheng and Qifeng Wu. A decision-making method for fire detection data fusion based on bayesian approach. In *2013 Fourth International Conference on Digital Manufacturing Automation*, pages 21–23, 2013.

[66] Lotfi.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[67] Zhu Jian, Cao Hongbing, Shen Jie, and Liu Haitao. Data fusion for magnetic sensor based on fuzzy logic theory. In *2011 Fourth International Conference on Intelligent Computation Technology and Automation*, volume 1, pages 87–92, 2011.

[68] Hamid Medjahed, Dan Istrate, Jerome Boudy, Jean-Louis Baldinger, and Bernadette Dorizzi. A pervasive multi-sensor data fusion for smart home healthcare monitoring. In *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, pages 1466–1473, 2011.

[69] Brandon Cook and Kelly Cohen. Multi-source sensor fusion for small unmanned aircraft systems using fuzzy logic. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6, 2017.

[70] E.H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1):1–13, 1975.

[71] Tomohiro Takagi and Michio Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-15(1):116–132, 1985.

[72] T. Denoeux. A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(5):804–813, 1995.

[73] Lotfi A. Zadeh. A simple view of the dempster-shafer theory of evidence and its implication for the rule of combination. *AI Mag.*, 7:85–90, 1985.

[74] Ronald R Yager and Liping Liu. *Classic Works of the Dempster-Shafer Theory of Belief Functions: An Introduction*, pages 1–34. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[75] Didier Dubois and Henri Prade. Representation and combination of uncertainty with belief functions and possibility measures. *ComputationalIntelligence*, 4, 1988.

[76] Philippe Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990.

[77] Philippe Smets. Belief functions: The disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, 9(1):1–35, 1993.

[78] Eric Lefevre, Olivier Colot, and Patrick Vannoorenberghe. Belief function combination and conflict management. *Information Fusion*, 3(2):149–162, 2002.

[79] Catherine K. Murphy. Combining belief functions when evidence conflicts. *Decision Support Systems*, 29(1):1–9, 2000.

[80] Deng Yong, Shi WenKang, Zhu ZhenFu, and Liu Qi. Combining belief functions based on distance of evidence. *Decision Support Systems*, 38(3):489–493, 2004.

[81] Zhenjiang Zhang, Tonghua Liu, Dong Chen, and Wenyu Zhang. Novel algorithm for identifying and fusing conflicting data in wireless sensor networks. *Sensors (Basel, Switzerland)*, 14:9562 – 9581, 2014.

[82] Cholsok Yu, Jianhong Yang, Debin Yang, Xianghong Ma, and Hyonchun Min. An improved conflicting evidence combination approach based on a new supporting probability distance. *Expert Systems with Applications*, 42(12):5139–5149, 2015.

[83] Tian Jing. An evidence fusion method using generalized mahalanobis distance in dempster-shafer theory. In *2015 4th International Conference on Computer Science and Network Technology (ICCSNT)*, volume 01, pages 470–473, 2015.

[84] Yibing Li, Jie Chen, Fang Ye, and Dandan Liu. The improvement of ds evidence theory and its application in ir/mmw target recognition. *J. Sensors*, 2016:1903792:1–1903792:15, 2015.

[85] Yuan Kaijuan, Xiao Fuyuan, Fei Liguo, Kang Bingyi, and Deng. Conflict management based on belief function entropy in sensor fusion. *SpringerPlus*, 5, 2016.

[86] Yong Deng. Deng entropy. *Chaos, Solitons Fractals*, 91:549–553, 2016.

[87] Fang Ye, Jie Chen, and Yibing Li. Improvement of ds evidence theory for multi-sensor conflicting information. *Symmetry*, 9:69, 2017.

[88] Yongchuan Tang, Deyun Zhou, Shuai Xu, and Zichang He. A weighted belief entropy-based uncertainty measure for multi-sensor data fusion. *Sensors (Basel, Switzerland)*, 17, 2017.

[89] Yun Lin, Can Wang, Chunguang Ma, Zheng Dou, and Xuefei Ma. A new combination method for multisensor conflict information. *The Journal of Supercomputing*, 72:2874–2890, 2016.

[90] Jiyao An, Meng Hu, Li Fu, and Jiawei Zhan. A novel fuzzy approach for combining uncertain conflict evidences in the dempster-shafer theory. *IEEE Access*, 7:7481–7501, 2019.

[91] Junwei Li, Baolin Xie, Yong Jin, Zhentao Hu, and Lin Zhou. Weighted conflict evidence combination method based on hellinger distance and the belief entropy. *IEEE Access*, 8:225507–225521, 2020.

[92] Zhan Deng and Jianyu Wang. A novel evidence conflict measurement for multi-sensor data fusion based on the evidence distance and evidence angle. *Sensors*, 20(2):381, Jan 2020.

[93] Hanwen Li and Fuyuan Xiao. A method for combining conflicting evidences with improved distance function and tsallis entropy. *International Journal of Intelligent Systems*, 35:1814 – 1830, 2020.

[94] Lifan Sun, Yuting Chang, Jiexin Pu, Haofang Yu, and Zhe Yang. A weighted evidence combination method based on the pignistic probability distance and deng entropy. *Journal of Aerospace Technology and Management*, 12:1–14, 2020.

[95] Hangyu Yan and Yong Deng. An improved belief entropy in evidence theory. *IEEE Access*, 8:57505–57516, 2020.

[96] Deyun Zhou, Yongchuan Tang, and Wen Jiang. A modified belief entropy in dempster-shafer framework. *PLoS ONE*, 12, 2017.

[97] Lei Chen, Ling Diao, and Jun Sang. A new method to handle conflict when combining evidences using entropy function and evidence angle with an effective application in fault diagnosis. *Mathematical Problems in Engineering*, 2020.

[98] Nimisha Ghosh, Sayantan Saha, and Rourab Paul. idcr: Improved dempster combination rule for multisensor fault diagnosis. *Engineering Applications of Artificial Intelligence*, 104:104369, 2021.

[99] Yongchuan Tang, Xueyi Fang, Deyun Zhou, and Xiaofeng Lv. Weighted deng entropy and its application in uncertainty measure. *2017 20th International Conference on Information Fusion (Fusion)*, pages 1–5, 2017.

[100] Fuyuan Xiao, Zehong Cao, and Alireza Jolfaei. A novel conflict measurement in decision-making and its application in fault diagnosis. *IEEE Transactions on Fuzzy Systems*, 29(1):186–197, 2021.

[101] Wen Jiang. A correlation coefficient of belief functions. *Int. J. Approx. Reason.*, 103:94–106, 2016.

[102] Chaosheng Zhu and Fuyuan Xiao. A belief hellinger distance for d–s evidence theory and its application in pattern recognition. *Engineering Applications of Artificial Intelligence*, 106:104452, 2021.

[103] Ihsan Ullah, Joosang Youn, and Youn-Hee Han. Multisensor data fusion based on modified belief entropy in dempster–shafer theory for smart environment. *IEEE Access*, 9:37813–37822, 2021.

[104] Kaiyi Zhao, Rutai Sun, Li Li, Manman Hou, Gang Yuan, and Ruizhi Sun. An improved evidence fusion algorithm in multi-sensor systems. *Applied Intelligence*, 51:7614 – 7624, 2021.

[105] Yu-Cui Wang, Jian Wang, Meng-Jie Huang, and Minghui Wang. An evidence combination rule based on a new weight assignment scheme. *Soft Computing*, 26:7123 – 7137, 2022.

[106] Mengmeng Ma and Jiyao An. Combination of evidence with different weighting factors: A novel probabilistic-based dissimilarity measure approach. *Journal of sensors*, 2015.

[107] Eric Lefevre, Olivier Colot, Patrick Vannoorenberghe, and D. de Brucq. A generic framework for resolving the conflict in the combination of belief structures. *Proceedings of the Third International Conference on Information Fusion*, 1, 2000.

[108] Fang Ye, Jie Chen, and Yuan Tian. A robust ds combination method based on evidence correction and conflict redistribution. *J. Sensors*, 2018:1–12, 2018.

[109] Claude E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:623–656, 1948.

[110] Ronald R. Yager. Entropy and specificity in a mathematical theory of evidence. *Int. J. Gen. Syst.*, 9(4):249–260, Jan 1983.

[111] Didier Dubois and Henri Prade. A note on measures of speciality for fuzzy sets. *International Journal of General Systems*, 10(4):279–283, 1985.

[112] 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, 1990.

[113] Yonggang Zhao, Duofa Ji, Xiaodong Yang, Liguo Fei, and Changhai Zhai. An improved belief entropy to measure uncertainty of basic probability assignments based on deng entropy and belief interval. *Entropy*, 21(11):1122, Nov 2019.

[114] Moïse Digrais Mambe, Souleymane Oumtanaga, and Georges Nogbou 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, 2018.

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

[116] Fabio Cuzzolin. A geometric approach to the theory of evidence. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(4):522–534, 2008.

[117] Tian Jing. An evidence fusion method using generalized mahalanobis distance in dempster-shafer theory. In *2015 4th International Conference on Computer Science and Network Technology (ICCSNT)*, volume 01, pages 470–473, 2015.

[118] Bjrnar Tessem. Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61(2):315–329, 1993.

[119] Yafei Song, Xiaodan Wang, Lei Lei, and Aijun Xue. Evidence combination based on credibility and separability. *2014 IEEE 12th International Conference on Signal Processing (ICSP)*, pages 1392–1396, 2014.

[120] Cuiping Cheng and Fuyuan Xiao. A new distance measure of belief function in evidence theory. *IEEE Access*, 7:68607–68617, 2019.

[121] Nour El Imane Hamda, Allel Hadjali, and Mohand Lagha. An advanced weighted evidence combination method for multisensor data fusion in iot. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, pages 810–815, 2022.

[122] Nour El Imane Hamda, Allel Hadjali, and Mohand Lagha. Multisensor data fusion in iot environments in dempster–shafer theory setting: An improved evidence distance-based approach. *Sensors*, 23(11), 2023.

[123] 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, 2016.

[124] Dan Wang, Jiale Gao, and Daijun Wei. A new belief entropy based on deng entropy. *Entropy*, 21(10):987, Oct 2019.

[125] Hongfei Wang, Xinyang Deng, Wen Jiang, and Jie Geng. A new belief divergence measure for dempster–shafer theory based on belief and plausibility function and its application in multi-source data fusion. *Engineering Applications of Artificial Intelligence*, 97:104030, 2021.

[126] Fuyuan Xiao. Complex pignistic transformation-based evidential distance for multisource information fusion of medical diagnosis in the iot. *Sensors*, 21(3):840, Jan 2021.

[127] Wen Jiang, Boya Wei, Xiyun Qin, Jun Zhan, and Yongchuan Tang. Sensor data fusion based on a new conflict measure. *Mathematical Problems in Engineering*, 2016:1–11, 2016.

[128] Zirui Liao, Shaoping Wang, Jian Shi, Zhiyong Sun, Yuwei Zhang, and Muhammad Baber Sial. Cooperative situational awareness of multi-uav system based on improved d-s evidence theory. *Aerospace Science and Technology*, 142:108605, 2023.