Electronic Government and Electronic Participation
H.J. Scholl et al. (Eds.)

doi:10.3233/978-1-61499-670-5-211

Dealing with Imperfect Data in "Smart-Cities"

Hatem BEN STA [a,b,c], Amal BEN REJEB [b,d] , Said GATTOUFI [b,e]

[a] *University of Tunis, Higher Institute of Management, SOIE lab*

[b] *University of Tunis at El Manar, Higher Institute of Computer Science of El Manar Tunis, Tunisia*

[c] hatem.bensta@gmail.com [d] benrajabamal.ihec@gmail.com [e] algattoufi@yahoo.com

Abstract. As a new form of sustainable development, the concept "Smart Cities" knows a large expansion during the recent years. It represents an urban model, refers to all alternative approaches to metropolitan ICTs case to enhance quality and performance of urban service for better interaction between citizens and government. However, the smart cities based on distributed and autonomous information infrastructure contains millions of information sources that will be expected more than 50 billion devices connected by using IoT or other similar technologies in 2020. Real-time data generated from autonomous and distributed sources can contain all sorts of imperfections regarding on the quality of data e.g. imprecision, uncertainty, ignorance and/or incompleteness. Any imperfection in data within smart city can have an adverse effect over the performance of urban services and decision making. In this context, we address in this article the problem of imperfection in smart city data. We will focus on handling imperfection during the process of information retrieval and data integration and we will create an evidential database by using the evidence theory in order to improve the efficiency of smart city.

Keywords. Smart Cities, ICT, Real-time data, Imperfection, Evidential database, Theory of belief functions, IoT, IoE, Crowdsourcing

1. Introduction

The emergence of Internet of Things (IoT) and Information and Communication Technology (ICT) promoted several concepts, "Smart City" is one of these concepts. It has been quite fashionable in the policy arena in the last few years [1] and holds today the world through its nature of research and its specific dimensions that include the *people*, *economy*, *mobility*, *natural environment*, *ICT infrastructure*, *lifestyle* and *public administration* [2]. This concept has been adopted since 2005 by a number of technology companies [3] (such as: Cisco, Microsoft, HP, IBM, Siemens, Oracle, etc). IBM described the smart city as "*one that makes optimal use of all the interconnected information available today to better understand and control its operations and optimize the use of limited resources* [4] and Cisco defined the smart cities as those who adopt "*scalable solutions that take advantage of information and communications technology to increase efficiencies, reduce costs and enhance quality of life* [5]. Therefore, the Smart Cities consist to use the ICT to be more intelligent and efficient in the use of resources in order to maximize the life quality of city's population. However, with a distributed and autonomous information infrastructure characterized by an open database, a distributed information system and an advanced technology, a particular attention was given to the validity and the reliability of the information circulated in smart cities. Several analytical criteria used to select the sources of information (such as: the reliability of the sources, the objectivity of the information, the exactitude of data). But, all these criteria are unable to estimate the reliability of the information sources. In fact, Real-time data generated from the different information sources can be for the most part, imprecise, uncertain, incomplete or ambiguous, which influences the efficiency of smart cities. In order to ensure a smart information infrastructure, we address in this paper the problem of imperfection in smart cities data. We model all the forms of imperfection by using the belief functions theory and we create evidential databases contains perfect and imperfect data where the imperfection is modeled with the Dempster-Shafer theory. In this context, we organize our article as follows: In section 2, we will draw a description of "Smart cities". In section 3, we will describe the problem of imperfection in smart city data. Section 4 will contain a description of our proposed method to deal with imperfect data and we will prove the steps of our approach in section 5. Finally, conclusion will draw.

2. Concept of "Smart Cities"

As a new form of sustainable development, the concept "Smart Cities" has attracted a lot of attention in the recent years [1]. Several definitions have been proposed to describe this concept. But, it still a vague or a fuzzy phenomenon [1], [6,7,8,9,10,11]. In this section, we aim to describe the Smart Cities and we aim to provide our own definition of this concept that we will hear a lot in the coming years.

2.1. Literature review: Definitions of Smart Cities

The definitions of smart cities are various and there are several researchers explored this area. Caragliu *et al.* believe that a city will be smart when the investments in human and social capital fuel a sustainable economy and a high quality of life, with a wise management of natural resources [1]. Harrison and Donnelly indicated in [3] that "*it's a new policy for urban planning*. [6] presented the smart cities by the utilization of ICT infrastructure, human resources, social capital and environmental resources in order to guarantee the economic development, the social sustainability and to ensure a high quality of human life. Vanolo considered the Smart city in [7] as an efficient city uses advanced technologies. Hollands mentioned in [8] that the smarter cities based on the utilization of network infrastructure to improve economic and political efficiency in order to guarantee the urban development. Ojo *et al.* described the smart cities in [9] as an urban innovation aim to harness physical and social infrastructures for economic regeneration, social cohesion and infrastructure management. Chourabi *et al.* indicated in [10] that "*the new intelligence of cities, resides in the increasingly effective combination of digital telecommunication networks (the nerves), ubiquitously embedded intelligence (the brains), sensors and tags (the sensory organs), and software (the knowledge and cognitive competence)*". Nam and Pardo [11] defined the concept of "Smart cities" as an "*organic connection among technological, human and institutional components*" and Schaffers *et al.* mentioned in [12] that it's a"*multi-dimensional concept. It is a future scenario, even more it an urban development strategy. It focuses on how technologies enhance the lives of citizens*". Generally, we can deduce through the current literature of Smart cities, two main definitions have been proposed to describe these cities. The first characterizes the smart cities by the wide use of ICT for traditional infrastructures for improving the active participation of human and social capital [1], [4,5,6,7,8]. The second defined the smart cities as the cities with smart physical, social and economic infrastructure while ensuring the centrality of citizens in a sustainable environment refer to the key characteristics defined by distinct factors (e.g., smart economy, smart mobility, smart people, smart environment, smart living, smart governance) and focus on the strategic use of new technology and innovative approaches to enhance the efficiencies and the competitiveness of cities [2], [9,10,11,12]. Therefore, we can define the concept "Smart City" as "*a modern city uses smart information infrastructure (contains perfect data) to ensure the sustainability and the competitiveness of the different urban functions by integrating different dimensions of urban development and investments in order to reduce the environmental impact and to improve the quality of citizens' lives*".

2.2. Smart Cities Applications

It all started in 2005 by several models of cities consists to implement complex information systems in urban infrastructure (such as buildings, transport, electricity, ...) in order to improve the quality of citizens' life. The first model of smart cities was proposed by Cisco in Dubai[1]. Cisco en-

[1]Cisco (2005): Smart City in Dubai. http://www.cisco.com/web/learning/le21/le34/downloads/689/nobel/2005/docs/Abdulhakim_Malik.pdf

ables Dubai a Smart Government (e-Government)[2], Smart Media City (DMC)[3], Healthcare City (DHC)[4] and Knowledge Village (DKV)[5]. Another model of Smart Cities was proposed by IBM in New York[6]. In this context, IBM provides set of applications, such as: The smarter transportation management network[7], Smarter Building Management[8], Smart water resources management[9], etc. Siemens, offers also a model of a smart city in Germany[10], as we can mention the model of smart city in Montreal[11]. Therefore, several models of smart cities have been proposed and all these models have the same components of *Smart-Economy*, *Smart-Mobility*, *Smart-Governance*, *Smart-Environment*, *Smart-Living* and *Smart-People* [13]. But, the integration of ICT in the different urban functions can pose certain problems, such as:

- Breach of confidentiality, the sensors monitor all the action of each individual, tracing... [6];
- Problem of restructuring [8];
- The emergence of new exclusion forms related to the inaccessibility of ICT and the reduction of creativity [14];
- The expensive installation of digital infrastructure [15];

We can conclude that the main challenge for smart city manifested essentially in its *information infrastructure* that is characterized by distributed and autonomous information sources generate large amount of imperfect data. This imperfection within smart city data can have an adverse effect over the performance of urban services and decision making. The following section describes the problem of imperfection and presents an actual example of imperfect data.

3. Smart Cities and imperfect data

Several objects, peoples, processes and devices communicate through internet-connected infrastructures in Smart Cities and generate a large amount of data, such as: the sensors, databases, media, etc. The emergence of ICT promotes several other information sources, such as: Cloud computing, IOT, Crowdsourcing, etc. Figure 1 summarizes the different information sources in smart cities.

However, the distributed and autonomous information infrastructure that specifies the smart cities poses several challenges related to the quality of data. Real-time data generated within this infrastructure can contain all sorts of imperfections in data (e.g. imprecision, uncertainty, ignorance, ambiguity, and/or incompleteness). For example, in the opinion individual's source like the "Crowdsourcing" that was popularized by Jeff Howe in 2006 [16] to execute the tasks that are hard for computers but easy for humans. The participants can answer by several solutions to a question which gives uncertain and/or imprecise response, they can skip to answer a question which give an incomplete or a missing response and they can answer by "I do not know" to reflect the ignorance [17]. All these types of imperfect information can have an adverse effect over the performance of urban services and decision making. Therefore, it's important to deal with

[2] www.dubai.ae
[3] www.dubaimediacity.com
[4] www.dhcc.ae
[5] www.kv.ae
[6] IBM (2009): Smarter Cities in New York 2009 http://www.ibm.com/smarterplanet/us/en/smarter_cities/article/newyork2009.html
[7] "Building a smarter transportation management network"
[8] "Smarter Buildings: Reduce cost and gain control"
[9] "Employing integrated operations for water resources management"
[10] "Pictures of the Future": http://www.siemens.com/innovation/en/home/pictures-of-the-future.html
[11] http://www.smartcityexpomtl.com/

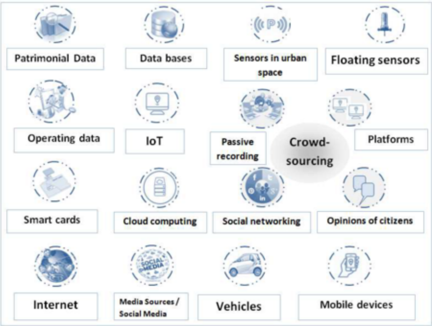

Figure 1. Typology of information sources in Smart Cities

Theories	Application areas	Source
Probability theory	Incompleteness	[26]
Fuzzy set logic	Imprecision and ambiguity	[28]
Possibility theory	Imprecision and uncertainty	[29]
Bipolar fuzzy sets	Non-existence information	[21]
Rough sets theory	Vagueness	[22]
Belief functions theory	Imprecision, uncertainty, incompleteness, ignorance and conflict	[23], [24]

Table 1. The uncertainty theories

all the forms of imperfection in order to improve the efficiency of smart cities. In this context, we focus in this paper on handling imperfect data during the process of information retrieval and data integration. The following section presents our approach to ensure perfect information infrastructure in Smart City.

4. Dealing with imperfect data in Smart Cities

To ensure the sustainability of the different urban functions, it must firstly guarantee an perfect information infrastructure. In the context of smart cities, there are several information coming from different sources, this information can be, for the most part, uncertain, imprecise, incomplete and/or missing. Several theories have been proposed to model data's imperfections such as: the probability theory [18] for modeling incomplete data, the possibility theory [19] for modeling imprecise data, the fuzzy set logic [20] for modeling ambiguity and imprecise data, we can also mention the bipolar logic [21] and the set approximate (Rough Sets) [22]. But, still the Dempster [23] Shafer [24] theory (DST) the most used theory. Its a mathematical theory represents a powerful tool enables to model all forms of imperfection (imprecision, uncertainty, ignorance, incompleteness and have access to conflict) [25]. Indeed, the Probability theory is the oldest theory for modeling incomplete data, but it cannot distinguish the uncertainty of the imprecision [26,27]. The fuzzy sets theory used only for modeling imprecision and vagueness [28]. Thus, Possibility theory offers a natural setting for representing only imprecise data and poor information [29]. However, the theory of belief function also referred to the evidence theory or DST provides a powerful tool for modeling all the kinds of imperfection. It's a flexible tool to take into account the imperfection of data in pattern recognition and information fusion. Table 1 summarizes the specificities of each theory to deal with the imperfection.

In this context, we resort to the belief functions theory in order to ensure a smart information infrastructure. According to this theory, we model all the forms of imperfection in smart city data and we create evidential databases containing both certain and/or uncertain data. We focus on handling the problem of imperfection in real-time data and provide mechanisms for real-time updates in evidential databases. The following sections present the basic concepts of Dempster-Shafer theory (section 4.1) and describe the D-S databases (section 4.2) and internet of everythings that it will be our application area (section 4.3).

4.1. Theory of belief functions (TBF)

Decision making is more difficult when handling imperfect information. Several theories have been proposed to model this imperfection. As they uncertainty theories like the theory of probability, the theory of possibility and the theory of fuzzy sets, the theory of belief functions models all the forms of imperfection. It's a mathematical theory represents a powerful tool for representing imperfect information. This theory was introduced firstly by Dempster [23] then formalized by Shafer [24]. The evidence theory gives a complete framework to model the imperfection in smart cities data. In this section, we introduce the fundamental notions of this theory and we present some related functions and some combination of rules that was later using to create the evidential databases (*EDB*).

4.1.1. Frame of discernment

A discernment frame $\Omega = \{\omega_1, \omega_2, \omega_3, ..., \omega_n\}$ is the set of all the exclusive and exhaustive hypotheses, called also the universe of discourse or domain of reference. The power set $2^\Omega = \{A|A \subseteq \Omega\} = \{\emptyset, \omega_1, \omega_2, \omega_3, ...,$
$\omega_n, \omega_1 \cup \omega_2, \Omega\}$ represents the set of all the hypothesis of Ω and their disjunctions.

4.1.2. Basic belief assignment (BBA)

A basic belief assignment or a mass function represents the degree of belief that supports the event (A). It affects a real value from [0 ; 1] and defined as follows:

$$m^\Omega : 2^\Omega \rightarrow [0,1] \tag{1}$$

$$\sum_{A \subseteq \Omega} m^\Omega(A) = 1 \tag{2}$$

We consider any positive elementary mass $m(A) > 0$ as a focal element such that A belongs to 2^Ω. If we have $m(\Omega) = 1$ that represent a total ignorance. If we consider a mass function such as $m(\{\omega_1, \omega_3\}) = 0.7$ and $m(\Omega) = 0.3$, this mass function model both imprecision (on $\{\omega_1, \omega_3\}$) and uncertainty with 0.7.

4.1.3. Particular belief functions

The mass function or the basic belief assignment represents a common representation of evidential knowledge, it has several categories and many particular functions. Such as:

Definition 1 *Categorical mass functions: A categorical BBA is a mass function noted by* m_A^Ω *which has a unique focal element* $A \subseteq \Omega : m_A^\Omega(A) = 1$.

Definition 2 *Vacuous mass functions: A vacuous BBA is a particular categorical mass function characterized by only one focal element A with* $A = \Omega$*, such that* $m^\Omega(A) = 1$. *This type of mass function is defined as follows:*

$$m^\Omega(A) = \begin{cases} 1 \; if A = \Omega \\ 0 \; otherwise \end{cases} \tag{3}$$

Definition 3 *Dogmatic mass functions: Dogmatic BBA characterized by a focal element different from Ω with $m(\Omega) = 0$.*

Definition 4 *Simple mass function: A simple BBA is a mass function which has only two focal elements.*

Definition 5 *Consonant mass functions: A consonant BBA is a mass function with the focal elements are nested, such as: $(A \subseteq B \subseteq ... \subseteq \Omega), \forall A, B \subseteq \Omega$ with $m(A) \neq 0$ and $m(B) \neq 0$.*

Definition 6 *Bayesian mass functions: A Bayesian BBA is a mass function which all the focal elements are singletons.*

4.1.4. The combination rules

There are several combination rules proposed in the context of belief functions. We start by the first combination rule that was proposed by [23]. For two mass functions m_1 and m_2 and $\forall X \in 2^{\Omega}$, the Dempsters combination rule ($m_{\oplus}$) is given by:

$$m_{1\oplus 2}(X) = m_1 \oplus m_2(X) = \frac{1}{1-K} \sum_{Y_1 \cap Y_2 = X} m_1(Y_1) m_2(Y_2) \quad (4)$$

Where $k = m_{\oplus}(\emptyset)$, and it's called the *global conflict*. In order to solve the problem enlightened by Zadeh's counter example [30] where the Dempster's rule produced unsatisfactory results, several combinations rules have been proposed. Smets improved in the Tranferable Belief Model [31] the Dempster's rule by *the conjunctive combination rule*. For two mass functions m_1 and m_2 and $\forall X \in 2^{\Omega}$, $m_{1 \bigcirc\!\!\!\!\cap\, 2}(A)$ is defined by:

$$m_{1 \bigcirc\!\!\!\!\cap\, 2}(A) = (m_1 \bigcirc\!\!\!\!\cap\, m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C) \quad (5)$$

4.2. Evidential database (EDB)

The databases used to store a large amount of information that can be uncertain or imprecise. To address this problem, the evidential databases have been proposed by Hewawasam *et al.* in [32] and Bach Tobji *et al.* in [33]. An evidential database is a database that contains perfect and imperfect data. Where the imperfection (uncertainty and / or imprecision) is represented by the belief functions theory with an evidential value V_{ij}. Formally, an evidential database is composed of X attributes (columns) and Y records (rows). Each attribute $j(1 < j < X)$ has a framework that represents all possible values of this attribute: This is the frame of discernment. The evidential value (V_{ij}) described by a mass function defined by:

$$m_{ij} : 2^{\Omega} : 2^{D_j} \rightarrow [0, 1] \quad (6)$$

$$m_{ij}(\emptyset) = 0 \sum_{x \subseteq D_j} m_{ij}(x) = 1 \quad (7)$$

4.3. Internet of everythings (IoE)

In smart cities all the objects, the people, the processes and the databases are connected to an Internet network. Internet of Everything is a networked connection of all of the information sources. This concept is a novel paradigm that is rapidly gaining ground in the scenario of modern wireless cities. Cisco was the founder of the concept of Internet of Everything (IoE) [34], it defined this concept as the brings of "*people, process, data and things to make networked connections more relevant and valuable than ever before turning information into actions that create new capabilities, richer experiences, and unprecedented economic opportunity for businesses, individuals, and countries*" [12]. Several models of Internet of Everything will be proposed in smart cities. Cisco was the leader on integrating Internet of Everything in Smart Cities [35] by a model

[12] "The Internet of Everything": Global Private Sector Economic Analysis

of IoE economics in Dubai (IoE To Drive Dubai's Smart Economy) [13] [14] [15]. Therefore, ensuring a reliable information infrastructure signified ensuring a reliable infrastructure for IoE in smart cities. In this context, we chose the environment of IoE to prove the importance of our approach in the the context of smart cities data.

5. Experimentation

In order to improve the efficiency of smart cities, we address the problem of handling imperfect data during the process of information retrieval and data integration. This imperfection manifested in the information circulated in the smart cities (Real-time data or data warehouse) can have several forms, such as:

- **Uncertain information**: It reflects the lack of knowledge (eg. "I think that the percent of water in the Earth's surface equal to 70%").
- **Imprecision information**: It translates the non-specificity (eg. "I think that the percent of water in the Earth's surface between 70% and 71%").
- **Vague information**: It reflects an ambiguous information (eg. "I think there are large amounts of water in Earth's surface).
- **Missing information**: It reflects the not found or incomplete information.

All these types of imperfect information influence the performance of urban services. Therefore, it's important to deal with the problem of imperfection to ensure a reliable information infrastructure. The following section presents the different steps of handling imperfect data with the evidence theory in smart city.

5.1. Experimental Setup

Handling imperfect data with the belief functions theory comprises two main steps: **representing data** and **modeling data**. In order to present the real knowledge and to improve the quality of real-time data, we estimate the reliability of the information sources and we integrate it in an evidential database. In this context, we will develop a platform based on the principles of IoE ensures the interconnection and the integration of the different information sources (objects, people, process and databases) and provides the opportunity to express the certainty level about the information. In this article we limited by modeling data coming from the opinion individual's source like "Crowdsourcing platforms". The following sections present the ways of representing and modeling data.

5.1.1. Presenting data

The main idea through the representation of data consists to deduct the imperfection that it will be modeled after with a mass function (*BBA*) and give the opportunity to present the uncertainty and the imprecision level. Generally, we assume that each data (D_i) coming from the source (s_j) is defined in the frame of discernment $\Omega_{s_j}^{D_i}$ and each frame belongs to a specific area (eg. transport, health, education, economy,...). Each information will have a degree of certainty D_c generated by the source of the information s_j and modeled after by a mass function $m_{s_j}^{D_i}$, which gives a matrix of I data/lines for J source/columns given by:

[13] "Dubai Smart City IoE Value at Stake in the Public Sector"
[14] "The Internet of Everything AED 17.9 bn Opportunity for Dubai:2014-2019"
[15] `http://www.gulfbusiness.com/articles/insights/internet-of-everything-to-drive-dubais-smart-economy/`

Type of data	Mass function (bba)	Case
1 Perfect data	$\{ m^{\Omega}(\omega_i) = 1 \}$	Perfect data
2 Ambiguous data		
2.1 Certain but imprecise data	$\{ m^{\Omega}(\omega_i \cup \omega_j) = 1 \}$	Possibilistic data
2.2 Precise but uncertain data	$\{m^{\Omega}(\omega_i) = 0.7\}, \{m^{\Omega}(\omega_j) = 0.3 \}$	Probabilistic data
3 Missing data	$\{ m^{\Omega}(\Omega) = 1 \}$	Total ignorance

Table 2. The cases of imperfection

$$\begin{array}{cc} & \begin{array}{ccccc} s_1 & \ldots & s_j & \ldots & s_J \end{array} \\ \begin{array}{c} D_1 \\ \vdots \\ D_i \\ \vdots \\ D_I \end{array} & \left[\begin{array}{ccccc|c} m_{s_1}^{\Omega_1} & \ldots & m_{s_j}^{\Omega_1} & \ldots & m_{s_J}^{\Omega_1} & \Omega_{s_j}^{D_1} \\ \vdots & & \vdots & & \vdots & \vdots \\ m_{s_1}^{\Omega_i} & \ldots & m_{s_j}^{\Omega_i} & \ldots & m_{s_J}^{\Omega_i} & \Omega_{s_j}^{D_i} \\ \vdots & & \vdots & & \vdots & \vdots \\ m_{s_1}^{\Omega_I} & \ldots & m_{s_j}^{\Omega_I} & \ldots & m_{s_J}^{\Omega_I} & \Omega_{s_j}^{D_I} \end{array}\right] \end{array} \tag{8}$$

The idea through the representation consists to better express the knowledge level of the source s_j about data with a certainty degree $D_c \in [0,...,1]$ will be modeled by a mass function $(m_{s_j}^{D_i})$ in order to present the imperfection level.

5.1.2. Modeling data

To model the received information, we assume that each data D_i proposed by the source s_j with $s_j = \{1,...,J\}$ belongs to a specific frame of discernment $\Omega_{s_j}^{D_i}$ with $\Omega_{s_j}^{D_i} = \{\omega_1, \omega_2, \omega_3, ..., \omega_n\}$. Where the power set $2^{\Omega} = \{A|A \subseteq \Omega\} = \{\emptyset, \omega_1, \omega_2, \omega_3, ..., \omega_n, \omega_1 \cup \omega_2, \Omega\}$ represents the set of all the hypothesis on Ω. The choice of the frame of discernment is extremely important to avoid the problem of complexity. For these reasons we limited the size of our frame of discernment between 2 and 6 focal elements, in order to guarantee an precise generation of the mass functions. Each focal element should be modeled by a mass function $(m_{s_j}^{D_i})$. The choice of *BBA* is done according to the categories of the selected focal element (ω_i). If the focal element (ω_i) is a singleton and its D_c equal to one, the bba will be a *certain bba* with $m_{s_j}^{D_i}(\omega_i) = 1$, which models the case of perfect information (precise and certain information), else if its $D_c \neq 1$ the bba will be a *bayesian bba* with $m_{s_j}^{D_i}(\omega_i) \in [0,...0.9]$. We are in the case of probabilistic information, which models the case of precise but uncertain information. When the focal element is Ω whith $m_{s_j}^{D_i}(\Omega) = 1$, we are in the case of the *total ignorance*. Finally, if the focal elements are nested $(\omega_1 \subseteq \omega_2 \subseteq \omega_3...)$, its bba will be a *consonant bba* with $m_{s_j}^{D_i}(\omega_1 \cup \omega_2 \cup \omega_3) \in [0,...,1]$. Table 2 summarizes the cases of modeling imperfect data with Dempster-Shafer theory.

5.1.3. Particular case: Handling imperfection in "Crowdsourcing platforms"

We present in this section a particular case of modeling imperfect data in "Crowdsourcing platforms" specific on healthcare area. The principle of "Crowdsourcing" consists to enlist a set of humans to solve some probem via the World-Wide Web. Ben Rjab *et al.* are already identified in [17] the reliable sources in crowdsourcing platforms with the evidence theory. In this context, we assume that there are only the experts in this platform. Therefore, the applicants ask the questions and the experts in health care should be respond by one or more answers. If the asked question (Q_i) was: *"What are the symptoms of Alzheimer's disease?"*. The frame of discernment of Q_i with $\Omega_{Q_i} = \{H_1, H_2, H_3, H_4\}$ will be:

- H_1: **Forgetfulness** with $D_c \in [0,..,1]$

- H_2: **Depression** with $D_c \in [0,..,1]$
- H_3: **Anger** with $D_c \in [0,..,1]$
- H_4: **Non discrimination** with $D_c \in [0,..,1]$

This algorithm (Algorithm 1) presents the steps to deduce the certainty degree. If an expert

---

Algorithm 1 CERTAINTY DEGREE D_c

Input:
I: Number of questions
J: Number of participants
Output:
D_c: Certainty degree
\# Initialization
know ← *False*
for $i \in [1:I]$ **do**
  for $j \in [1:J]$ **do**
    $Res[i][j]$ ← Response to a question
    if *(know = True)* **then**
      D_c ← An evidential value between [0,...,1]
    $D_c \leftarrow 0$
return D_c

(s_1) responds with a singleton focal element eg. $\{H_1\}$ with a certainty degree (D_c) equal to one. We are in the case of perfect response (certain and precise answer), a *certain bba* with $m_{s_1}^{\Omega_{Q_i}}(H_1) = 1$ will be added to this information. If the focal elements are singletons, but with a $D_c \neq 1$. We are in the case of precise but uncertain answer, a *bayesian bba* will be added to this information. If an expert (s_2) responds by $\{H_1 \cup H_2\}$ with a degree of belief (D_c) equal to 1. We are in the case of imprecise (on $\{H_1, H_2\}$) but certain answer, a *consonant bba* with $m_{s_3}^{\Omega_{Q_i}}(H_1 \cup H_2) = 1$ will be added to this information. But, if an expert (s_3) responds by $\{H_1 \cup H_2 \cup H_3\}$ with a $D_c \neq 1$ eg. 0.7. In this case, we have the uncertainty on the belief degree of 0.7 and the imprecision on $\{H_1$, H_2, $H_3\}$. Finally, if an expert (s_4) respond with $\{H_1, H_2, H_3, H_4\}$, in this case a *Vacuous bba* will be added to this information with $m_{s_4}^{\Omega_{Q_i}}(\Omega) = 1$ which reflects the total ignorance. Therefore, we can obtain for each question a matrix as follows:

5.2. *Experimental Results*

The result of our work manifested in an *evidential database (EDB)* also called *D-S database* includes all the perfect and imperfect data coming from the different sources. The imperfection in the evidential databases are expressed with the theory of belief functions presented above. Table 4 present an example of evidential table in evidential database stores perfect and imperfect data coming from the different participants (the experts) in "Crowdsourcing platforms" of health care area, where the imperfection is modeled by an evidential value V_{ij}.

As we have already explained, if the focal element is a singleton and its mass function equal to one, its bba will be a *"certain bba"*. We are in the case of *perfect information*, else if the focal elements are singletons and its $D_c \neq 1$, we are in the case of *probabilistic information*, its bba will be a *"bayesian bba"*. Else if, the focal element is Ω where its mass function is equal to one, its bba will be a *"Vacuous bba"* and we are in the case of the *"total ignorance"*. Else, we are the case of *possibilistic information* with *consonant bba*. Therefore, integrating evidential databases in smart cities promotes the sustainability of the different urban functions, improves the decision making and the efficiency of smart cities.

$[[Q_i]]$	s_1	s_j	s_J
H_1	$m_{s_1}^{\Omega_{Q_i}}(H_1)$	$m_{s_j}^{\Omega_{Q_i}}(H_1)$	$m_{s_J}^{\Omega_{Q_i}}(H_1)$
H_2	$m_{s_1}^{\Omega_{Q_i}}(H_2)$	$m_{s_j}^{\Omega_{Q_i}}(H_2)$	$m_{s_J}^{\Omega_{Q_i}}(H_2)$
$H_1 \cup H_2$	$m_{s_1}^{\Omega_{Q_i}}(H_1 \cup H_2)$	$m_{s_j}^{\Omega_{Q_i}}(H_1 \cup H_2)$	$m_{s_J}^{\Omega_{Q_i}}(H_1 \cup H_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
H_n	$m_{s_1}^{\Omega_{Q_i}}(H_n)$	$m_{s_j}^{\Omega_{Q_i}}(H_n)$	$m_{s_J}^{\Omega_{Q_i}}(H_n)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Ω	$m_{s_1}^{\Omega_{Q_i}}(\Omega)$	$m_{s_j}^{\Omega_{Q_i}}(\Omega)$	$m_{s_J}^{\Omega_{Q_i}}(\Omega)$

Table 3. Mass functions coming from the different sources

To combine the different opinions offered by the different participants in the crowd, there are several combination rules expressed via the evidence theory. In this context, we use the *conjunctive combination rule* that it was proposed by Smets improved in the Tranferable Belief Model [31]. We chose this combination rule because the sources in this platform are reliable. The combination of the different mass functions ($m_{s_j}^{\Omega_{Q_i}}$) generated by the sources (s_j) is very important to implement the evidential database that will be illustrated in the next section.

ID	Symptoms of Alzheimer disease	Evidential value (V_{ij})
1	$\{Forgetfulness\}$	$m_{s_J}^{\Omega_{Q_i}}(H_1) = 1$
2	$\{Forgetfulness \cup Depression\}$	$m_{s_J}^{\Omega_{Q_i}}(H_1 \cup H_2) = 0.7$ $m_{s_J}^{\Omega_{Q_i}}(\Omega) = 0.3$
3	$\{Depression\}, \{Anger\}$	$m_{s_J}^{\Omega_{Q_i}}(H_2) = 0.6$ $m_{s_J}^{\Omega_{Q_i}}(H_3) = 0.4$
4	$\{Forgetfulness \cup Depression \cup Anger \cup Non-discrimination\}$	$m_{s_J}^{\Omega_{Q_i}}(\Omega) = 1$

Table 4. Example of evidential table

6. Conclusion

With growing popularity of IoT and sensor technologies a large amount of data will be produced by different devices in the context of smart cities. Analyzing real-time data and handling imperfect information represent the main challenges of smart cities. In this context, we focus on dealing imperfection in smart cities data. We limited in this article by modeling data coming from the individual's source. We offered the opportunity for the individuals to express their certainty level about the added information, we modeled the data with the basic concepts of the belief functions theory and we integrated it in evidential databases by using such combination rule. Modeling imperfect data and integrating it in evidential databases promote the urban development, improve the decision making and increase the efficiency of smart cities. The use of the different concepts modeled in the evidential databases in such semantic models (ontologies) guarantees an evidential data interoperability in smart city. In another paper we will show how modeling imperfect data coming from other information sources (like the objects, processors, databases) and we will present how integrate it in evidential databases in order to ensure a reliable information infrastructure for smart cities.

References

[1] Andrea Caragliu, Chiara Del Bo, and Peter Nijkamp. Smart cities in europe. *Journal of urban technology*, 18(2):65–82, 2011.

[2] Adegboyega Ojo, Zamira Dzhusupova, and Edward Curry. Exploring the nature of the smart cities research landscape. In *Smarter as the New Urban Agenda*, pages 23–47. Springer, 2016.

[3] Colin Harrison and Ian Abbott Donnelly. A theory of smart cities. In *Proceedings of the 55th Annual Meeting of the ISSS-2011, Hull, UK*, volume 55, 2011.

[4] Michael Kehoe, Michael Cosgrove, S Gennaro, Colin Harrison, Wim Harthoorn, John Hogan, John Meegan, Pam Nesbitt, and Christina Peters. Smarter cities series: A foundation for understanding ibm smarter cities. *Redguides for Business Leaders, IBM*, 2011.

[5] Gordon Falconer and Shane Mitchell. Smart city framework: a systematic process for enabling smart+ connected communities. *Cisco internet business solutions group. Available online at: http://www. cisco. com/web/about/ac79/docs/ps/motm/Smart-City-Framework. pdf [Accessed 5 September 2013]*, 2012.

[6] SMART CITIES. Trace analysis and mining for smart cities: issues, methods, and applications. *IEEE Communications Magazine*, 121, 2013.

[7] Alberto Vanolo. Smartmentality: The smart city as disciplinary strategy. *Urban Studies*, page 0042098013494427, 2013.

[8] Robert G Hollands. Will the real smart city please stand up? intelligent, progressive or entrepreneurial? *City*, 12(3):303–320, 2008.

[9] Adegboyega Ojo, Edward Curry, and Tomasz Janowski. Designing next generation smart city initiatives-harnessing findings and lessons from a study of ten smart city programs. 2014.

[10] Hafedh Chourabi, Taewoo Nam, Shawn Walker, J Ramon Gil-Garcia, Sehl Mellouli, Karine Nahon, Theresa Pardo, Hans Jochen Scholl, et al. Understanding smart cities: An integrative framework. In *System Science (HICSS), 2012 45th Hawaii International Conference on*, pages 2289–2297. IEEE, 2012.

[11] Taewoo Nam and Theresa A Pardo. Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, pages 282–291. ACM, 2011.

[12] Hans Schaffers, Nicos Komninos, Marc Pallot, Miguel Aguas, Esteve Almirall, Tuba Bakici, Jean Barroca, Dave Carter, Michel Corriou, Joana Fernadez, et al. Smart cities as innovation ecosystems sustained by the future internet. 2012.

[13] Rudolf Giffinger, Christian Fertner, Hans Kramar, Robert Kalasek, Nataša Pichler-Milanovic, and Evert Meijers. Smart cities-ranking of european medium-sized cities. Technical report, Vienna University of Technology, 2007.

[14] Mike Davis. Who will build the ark. *New Left Review*, 61(2010):29–46, 2010.

[15] Leonidas Anthopoulos and Panos Fitsilis. From digital to ubiquitous cities: Defining a common architecture for urban development. In *Intelligent Environments (IE), 2010 Sixth International Conference on*, pages 301–306. IEEE, 2010.

[16] Jeff Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.

[17] A Ben Rjab, M Kharoune, Z Miklos, A Martin, and B Ben Yaghlane. Caractérisation dexperts dans les plate-formes de crowdsourcing characterization of experts in crowdsourcing platforms. *24 me Conference sur la Logique Floue et ses Applications (LFA), Poitiers, France*, (13), 2015.

[18] William Feller. An introduction to probability theory and its applications. vol. i. 1950.

[19] Didier Dubois and Henri Prade. *Possibility theory*. Wiley Online Library, 1988.

[20] Lotfi A Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy sets and systems*, 90(2):111–127, 1997.

[21] Horst H Berger and Siegfried K Wiedmann. Merged-transistor logic (mtl)-a low-cost bipolar logic concept. *Solid-State Circuits, IEEE Journal of*, 7(5):340–346, 1972.

[22] Zdzisław Pawlak. Rough sets. *International Journal of Computer & Information Sciences*, 11(5):341–356, 1982.

[23] Arthur P Dempster. Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics*, pages 325–339, 1967.

[24] Glenn Shafer et al. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.

[25] Isabelle Bloch. *Information fusion in signal and image processing: major probabilistic and non-probabilistic numerical approaches*. John Wiley & Sons, 2013.

[26] Didier Dubois, Helene Fargier, and Philippe Fortemps. Fuzzy scheduling: Modelling flexible constraints vs. coping with incomplete knowledge. *European Journal of Operational Research*, 147(2):231–252, 2003.

[27] Robert B Ash. *Basic probability theory*. Courier Corporation, 2012.

[28] Philippe Smets. Imperfect information: Imprecision and uncertainty. In *Uncertainty Management in Information Systems*, pages 225–254. Springer, 1997.

[29] Didier Dubois and Henri Prade. Possibility theory and data fusion in poorly informed environments. *Control Engineering Practice*, 2(5):811–823, 1994.

[30] Lotfi A Zadeh. Review of a mathematical theory of evidence. *AI magazine*, 5(3):81, 1984.

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

[32] KKRGK Hewawasam, K Premaratne, SP Subasingha, and M-L Shyu. Rule mining and classification in imperfect databases. In *Information Fusion, 2005 8th International Conference on*, volume 1, pages 8–pp. IEEE, 2005.

[33] Mohamed Anis Bach Tobji, Boutheina Ben Yaghlane, and Khaled Mellouli. A new algorithm for mining frequent itemsets from evidential databases. In *Proceedings of IPMU*, volume 8, pages 1535–1542, 2008.

[34] Dave Evans. The internet of things: How the next evolution of the internet is changing everything. *CISCO white paper*, 1:1–11, 2011.

[35] S Mitchell, N Villa, M Stewart-Weeks, and A Lange. The internet of everything for cities: connecting people, process, data and things to improve the livability of cities and communities, 2013.