2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress

Efficient Processing of Uncertain Data Using Dezert-Smarandache Theory: A Case Study

Hossein Jafari, Xiangfang Li, Lijun Qian
Center of Excellence in Research and Education for Big Military Data Intelligence (CREDIT)
Prairie View A&M University, Texas A&M University System
Prairie View, Texas 77446, USA
Email: hjafari@student.pvamu.edu, xili@pvamu.edu, lijunqian@ieee.org

***Abstract*—Dezert-Smarandache Theory (DSmT) of plausible and paradoxical reasoning has excellent performance when the data contain uncertainty or conflicting. However, the methods developed in DSmT are in general very computationally expensive, thus they may not be directly applied to multiple data sources with high cardinality. In this paper, we explore the feasibility of using DSmT in practical applications through a case study. Specifically, we propose a DSm hybrid model with reduced number of classes and thus low computational cost to analyze temperature and humidity data received from multiple sensors to determine comfort zones in a smart building. Data from each sensor is considered as individual evidence that can be uncertain, imprecise and even conflicting. Several types of combination rules are applied to calculate the total mass function. Then the belief, plausibility and pignistic probability are deduced. They are used as metrics for decision making to determine comfort levels of the monitored environment. Both simulation and real data experiments demonstrate that the proposed method would make DSmT feasible for practical situation awareness applications.**

***Keywords*-Dezert-Smarandache Theory (DSmT), Dempster-Shafer Theory (DST), Comfort Zone, Uncertain Data Fusion, Smart Building, Multi Sensor, Multi Hypothesis.**

I. Introduction

In future smart buildings or smart environment, numerous sensors will be deployed for monitoring and surveillance. As a result, large amount of data will be collected from various sources. In many practical cases, the data may contain uncertainties and sometimes even are conflicting. How to use the data to make inference and decisions becomes a challenge.

Dempster-Shafer theory [1] has been used to combine data (called evidence) from multiple sources. Compared to traditional Bayesian method, Dempster-Shafer theory has more flexibility in specifying ignorance and uncertainty in the data. When conflicts level among source of data become large and the refinement of frame of discernment is inaccessible because of the vague and imprecise nature of elements of frame of discernment, Dezert-Smarandache theory of plausible and paradoxical reasoning (DSmT) [2] can be applied as a powerful tool to combine the data. However, the methods in the DSmT framework are in general very computationally expensive, thus in many big data processing, they may not be directly applied to multiple data sources with high cardinality.

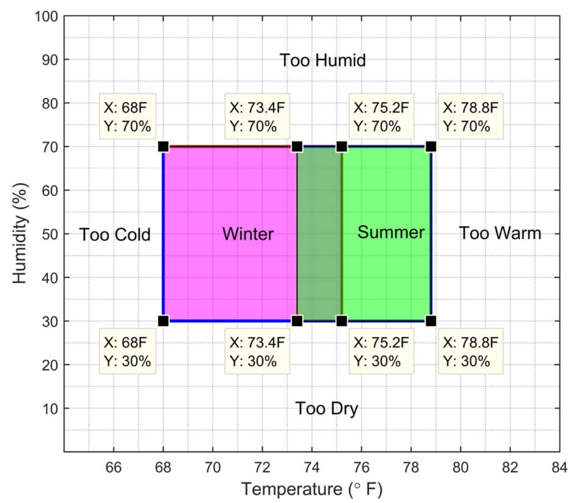

Figure 1. Relative humidity / temperature comfort zone (ISO7730-1984)

In this paper, we explore the feasibility of using DSmT in practical applications through a case study. Specifically, we propose a modified algorithm to use DSmT with reduced computational cost to analyze temperature and humidity data received from multiple sensors to determine comfort zones in a smart building. Comfort zone is defined as the range of temperature and humidity that people are feeling comfortable [3]. It is known as a thermal/human comfort too. Evaluating comfort zone is related to different parameters and even different from person to person. Fig.1 shows the “Comfort Zone” according to ISO7730-1984 standard. It designed based on several experiments and a large amount of empirical data that collected over several years from different locations. As these graphs display, comfort zone is different for winter and summer seasons.

In traditional buildings, the sensors are installed in some fixed places and they may not be able to measure locations of interest. In our previous work [4], we proposed a novel

DOI 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.127

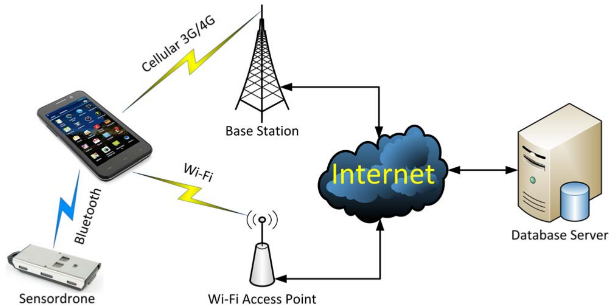

Figure 2. The architecture of the proposed community sensing system

framework of an environment air quality monitoring system based on community sensing, see Fig.2. Leveraging on the high penetration of smartphones and low cost and small form factor of certain sensors with a Bluetooth module, critical measurements such as air quality can be measured by each sensor carried by a member of a community, and be sent to that person's smartphones, and eventually uploaded to server using a corresponding app. Then the aggregated data at the server side can be processed to determine comfort zone and control HVAC [1] system to optimize the usage of electricity, while keeping the inhabitants comfortable. In this project, we have designed the architecture of the proposed community sensing system, and implemented the system using commercial off-the-shelf (COTS) Sensordrone [5], paired with Android© smartphones. Our system measures temperature, humidity, pressure, carbon monoxide, and battery charge level in real-time and it provided the experimental data in this study.

In this paper, we start by introducing the details of Dempster-Shafer theory of evidence (DST) in Section II and Dezert-Smarandache theory (DSmT) of plausible and paradoxical reasoning in Section III. Then we propose our models and apply different combination rules to calculate total mass, belief, plausibility and pignistic probability. Finally decision making based on those metrics are used to compare for different models and combination rules. Section IV describes data collection as our real data evidences. Section V explains our case studies including both synthetic data and real data analysis using the proposed methods with several types of combination rules. Section VI concludes the paper.

II. Dempster-Shafer Theory

Dempster-Shafer theory (DST) of evidence, or DST, is firstly originated by Dempster's work [6] on the upper and lower probabilities and later extended by Shafer's work [1] on the belief functions. It is an extension of the traditional Bayesian probability that gives capability to deal with uncertainty. To better understand Dempster-Shafer theory, we firstly introduce some propositions [7]:

[1]Heating, ventilation, and air conditioning

Frame of discernment: let Θ be a finite set of elements. Elements here refer to hypothesis or classes that for our study are feeling zones. Θ called the frame of discernment. For Dempster-Shafer model, all elements of Θ are assumed be exclusive and exhaustive. The power set of Θ that includes all subset of Θ is defined by 2^Θ. Basically power set includes all the elements of Θ and all combinations of their union. So it is closed under union operator.

Mass Function: mass function or basic belief assignment (bba) m is defined as a probability function. It maps a number in [0,1] to elements of 2^Θ in such a way that:

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

$$m(\emptyset) = 0 \tag{2}$$

$$\sum_{A \subseteq 2^\Theta} m(A) = 1 \tag{3}$$

Here $m(A)$ refers to the level of confidence in A, where A is a subset of 2^Θ. In our study, mass function refers to degree of belief for each class of feeling. In the case $m(A) > 0$, subset A is called a focal element. For the case subset A includes more than one element, because we do not have more information about each element separately, related mass function $m(A)$ cannot be decomposed to more mass functions for each individual element. One of the main differences between traditional Bayesian probability and Dempster-Shafer theory is the uncertainty function $m(\Theta)$ in DST:

$$m(\Theta) = 1 - \sum_{A \subset 2^\Theta} m(A) \tag{4}$$

Combination rule of Dempster-Shafer: In many big data applications, different types of data are aggregated from multiple sensors that may originated from multiple sources. Combined mass function can be calculated based on the Dempster-Shafer rule:

$$m(A) = m_1 \oplus m_2 \oplus ... \oplus m_n \tag{5}$$

$$m(A) = \begin{cases} 0 & A = \emptyset \\ \frac{\sum_{\bigcap_{k=1}^{n} A_k = A} m_1(A_1) m_2(A_2) ... m_n(A_n)}{1-K} & A \neq \emptyset \end{cases} \tag{6}$$

$$K = \sum_{\bigcap_{k=1}^{n} A_k = \emptyset} m_1(A_1) m_2(A_2) ... m_n(A_n) \tag{7}$$

$$1 - K = \sum_{\bigcap_{k=1}^{n} A_k \neq \emptyset} m_1(A_1) m_2(A_2) ... m_n(A_n) \tag{8}$$

Here K is the conflict value among all the sources of information. It is used as a normalization factor, $K \in (0,1)$. The higher value of K indicates more conflicting among information sources. As an example, for two sensors, Dempster-Shafer combinational rule is:

$$m(A) = m_1 \oplus m_2 \tag{9}$$

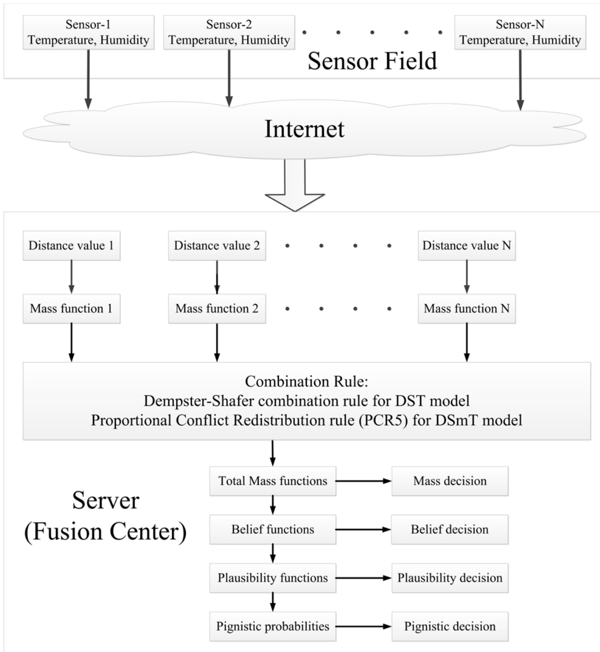

Figure 3. The proposed DST and DSmT combination and decision making for multiple data sources

$$m(A) = \begin{cases} 0 & A = \emptyset \\ \frac{\sum_{A_1 \bigcap A_2 = A} m_1(A_1) m_2(A_2)}{1-K} & A \neq \emptyset \end{cases} \tag{10}$$

$$K = \sum_{A_1 \bigcap A_2 = \emptyset} m_1(A_1) m_2(A_2) \tag{11}$$

$$1 - K = \sum_{A_1 \bigcap A_2 \neq \emptyset} m_1(A_1) m_2(A_2) \tag{12}$$

DST combination rule is associative, commutative and markovian. For the cases with more than two sources of data (called evidences in DST), DST combination rule can be extended by applying combination rule between two mass functions and then combine the result with new evidences and so on to compute combination for all sources of evidences. For DST combination we applied this method.

Associated with mass function, the belief function is defined as:

$$Bel(x) = \sum_{y \in 2^\Theta, y \subseteq x} m(y) \tag{13}$$

And plausibility function calculate as:

$$Pl(x) = \sum_{y \in 2^\Theta, x \bigcap y \neq \emptyset} m(y) = 1 - Bel(\bar{x}) \tag{14}$$

where $\bar{x}$ is the complement set of x, $\bar{x} = \Theta - x$. It is clear that $Pl(A) \geq Bel(A)$. Belief interval, $[Bel(A), Pl(A)]$, refers to the imprecision on the true probability, when belief function is the lower probability and plausibility function as an upper probability.

The pignistic probability introduced by [8] is defined as:

$$betP(x) = \sum_{y \in 2^\Theta, y \neq \emptyset} \frac{|x \cap y|}{|y|}.m(y) \tag{15}$$

where $|x|$ is the cardinality of x. Pignistic probability maps belief to probability to make a hard decision. As a result, belief functions provide a pessimistic view while plausibility function is optimistic. Pignistic probability is a compromise.

Reliable decision making using big data fusion is a challenge. Although there is not any unique metric for best decision making, four different metrics including total mass function, belief, plausibility and pignistic probability are tested in our simulation and experiment.

III. Dezert-Smarandache Theory

Dezert-Smarandache theory of plausible and paradoxical reasoning (DSmT) is an extension of DST and a generalized version of both DST and traditional Bayesian probability. DSmT has better performance when the uncertainty or conflicts among evidences are high. In DSmT, hyper power set of Θ is defined by D^Θ. It includes all the elements of Θ and all combinations of their union and intersection. Thus DSmT is closed under both union and intersection operators, while DST is closed under union operator only. Unlike DST, in DSmT we are not limited for exclusivity among elements of Θ. It is clear that the cardinality of hyper power set is much more than power set. Similar to DST, in DSmT mass function or generalized basic belief assignment (gbba) is defined as a mapping $m : D^\Theta \to [0, 1]$, $m(\emptyset) = 0$ and $\sum_{A \subseteq D^\Theta} m(A) = 1$. Belief, plausibility and generalized pignistic probability functions are defined as [2]:

$$Bel(x) = \sum_{y \in D^\Theta, y \subseteq x} m(y) \tag{16}$$

$$Pl(x) = \sum_{y \in D^\Theta, x \bigcap y \neq \emptyset} m(y) = 1 - Bel(\bar{x}) \tag{17}$$

$$betP(x) = \sum_{y \in D^\Theta, y \neq \emptyset} \frac{|\mathcal{C}_\mathcal{M}(x \cap y)|}{|\mathcal{C}_\mathcal{M}(y)|}.m(y) \tag{18}$$

where $\mid \mathcal{C}_\mathcal{M}(x) \mid$ is the cardinality, i.e., the number of parts x has in the model (Venn diagram).

Several combination rules have been developed based on DSmT model [2]. Those rules can manage or redistribute conflict values in different ways and have different complexity of computation. There are numerous combination rules can be defined to redistribute conflict values among elements of hyper power set. Classic DSm rule of combination, hybrid DSm rule, and series of proportional conflict redistribution rules (PCR) from PCR1 to PCR6 are some of those combination rules [2]. PCR5 is one of the most accurate rules in managing conflict. It redistributes partial conflict values just between the two elements that involved in that partial conflict. However, comparing to other methods it is hard to

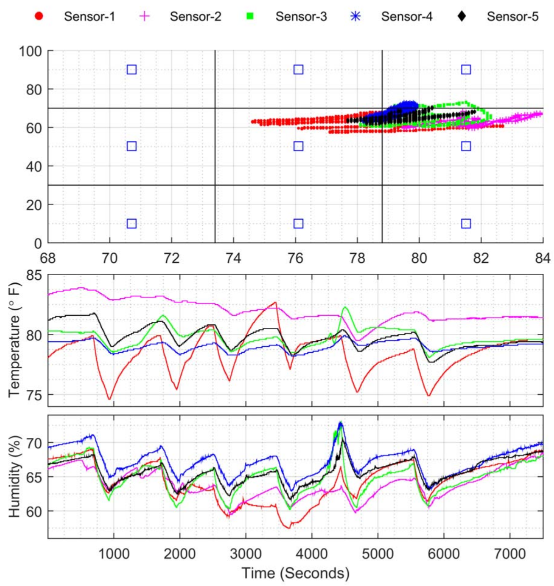

Figure 4. Temperature, Humidity data and Comfort zone

implement due to high computational cost. For two sources of evidences: $\forall X \in D^{\Theta} \setminus \{\emptyset\}$

$$m_{PCR5}(X) = m_{12}(X) + \sum_{Y \in D^{\Theta}, X \bigcap Y = \emptyset} \left[\frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right]$$

where m_{12} refers to conjunctive consensus:

$$m_{12}(X) = \sum_{X_1, X_2 \in D^{\Theta}, X_1 \bigcap X_2 = X} m_1(X_1) m_2(X_2) \quad (19)$$

PCR5 can be applied to more than two data sources [9]. Fig.3 shows the flowchart of applying DSmT combination rule (PCR5 as an example here) from sensing data to decision making.

Except the classic DSm rule of combination, all other combination rules based on DSmT model are non-associative and non-markovian. This implies that for more than two sources of evidences, combination rule cannot be applied blindly between two mass functions in repetitive way that we do in DST. Because these combination rules are non-associative and non-markovian, the order of sources in combination can change the result of combination. For calculating PCR5 rule for more than two sources we adopt a new method introduced in [10] to conserve the associativity and markovian property requirement to guarantee the correctness of the final combination. In fact, applying this algorithm transfers a non-associative and non-markovian rule to a quasi-associative and quasi-markovian rule.

To implement PCR5 rule of combination based on this algorithm for $n \geq 3$ sources, we firstly calculate conjunctive rule part, $m_{12}(X)$, between first two sources and transfer the whole conflict mass to empty or non empty set (we used non empty set Θ) and save the result. Then we calculate conjunctive rule between the saved results with the third source. We repeat this for first $n-1$ sources. Finally we apply PCR5 rule between the conjunctive result among $n-1$ sources and the last source. This algorithm has the advantage that the order of sources in the combination rule is no longer important and both associative and markovian properties are satisfied as well.

IV. Data Collection

In our experimental data collection, we used the proposed platform in Fig.2 to monitor the air quality inside an apartment during summer season. We put five sets of Sensordrones nodes and Android© smartphones with related apps in different parts of the apartment named room1, room2, living, dining, and kitchen, respectively. For all five sensor nodes, sensing interval is set to five seconds. The sensor nodes measured temperature, humidity, pressure, carbon monoxide, and battery level of sensor node. In addition, time stamps and GPS location data are uploaded to the server. We only used temperature and humidity data for our case study.

As an example, the monitored temperature and humidity data for a ten-hour sensing period including 7500 data samples and their mappings in comfort zone are shown in Fig.4. Fluctuations in temperature and humidity are caused by running of air conditioner (AC) periodically for cooling during the summer. AC was set to 77 degree Fahrenheit. AC was turned off for the last four hours. Then temperature and humidity started to increase in all places as expected.

V. Data Analysis and Decision Making

This section explains the details of our proposed model and implementation of DST and DSmT related combination rules based on our model. Both simulation results and real data analysis based on the experiments are shown to determine the comfort zone inside the apartment. According to the “Comfort Zone” in ISO7730-1984 standard [3] shown in Fig.1, we defined 9 classes/zones including the comfort zone and 8 other classes around the comfort zone. We will call this model the first model. Fig.5 shows the 9 classes for the summer season. In Fig.5, small blue square markers show the center of related classes and red solid lines are used as a boundary to distinguish between different classes. Table.I displays temperature and humidity values and feeling definition for related classes based on the first model. Thus the frame of discernment for feeling zone evaluation is

$$\Theta = \{l_1 = \text{“}I\text{”}, l_2 = \text{“}II\text{”}, ..., l_9 = \text{“}IX\text{”}\} \quad (20)$$

Here “I” refers to the first class and “II” refers to the second class and so on. Because all 9 classes are exclusive and

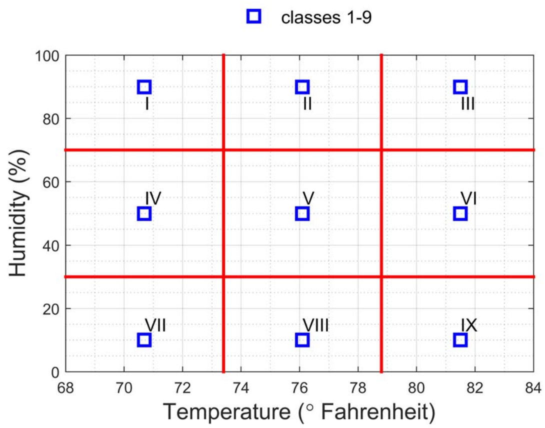

Figure 5. The 9 classes for summer season - First model

Table I
THE PROPOSED 9 CLASSES

Class	Temperature	Humidity	Feeling Definition
I	70.7	90	Too Cold & Humid
II	76.1	90	Too Humid
III	81.5	90	Too Warm & Humid
IV	70.7	50	Too Cold
V	76.1	50	Comfort Zone
VI	81.5	50	Too Warm
VII	70.7	10	Too Cold & Dry
VIII	76.1	10	Too Dry
IX	81.5	10	Too Warm & Dry

exhaustive, it satisfies the Dempster-Shafer model. It is noted that uncertainty is *not* considered in this model.

Feeling definitions in Table.I based on Fig.5 explain the human feeling for the range of temperature and humidity in each classes. For example, class “I” means “too cold and humid” and so on. Based on our proposed method in Fig.5, Dempster-Shafer combination rule can be applied to our data to compute total mass. Because in this model all 9 classes are singleton and exclusive, the total mass and belief functions are equal.

In order to calculate the mass function, we first calculate the normalized Euclidean distance between measured data from sensors and class parameters:

$$d_i^{l_j} = \left(\sum_{x=1}^{m} \left(\frac{S_x - f_x^{l_j}}{f_{max} - f_{min}} \right)^2 \right)^{1/2} \tag{21}$$

Here $d_i^{l_j}$ refers to the distance between sensor i and class j, S_x is sensor data and $f_x^{l_j}$ is the value of class j. $f_{max} - f_{min}$ is used for normalization. Then for sensor node k, distance values for all classes can be obtained:

$$D_k = \{d_k^{l_1}, d_k^{l_2}, ..., d_k^{l_n}\} \tag{22}$$

For the small value of distance $d_k^{l_i}$, the probability that the class of sensor k is l_i is higher. Then mass function can be calculated based on the distance values for each sensor node k and each class j:

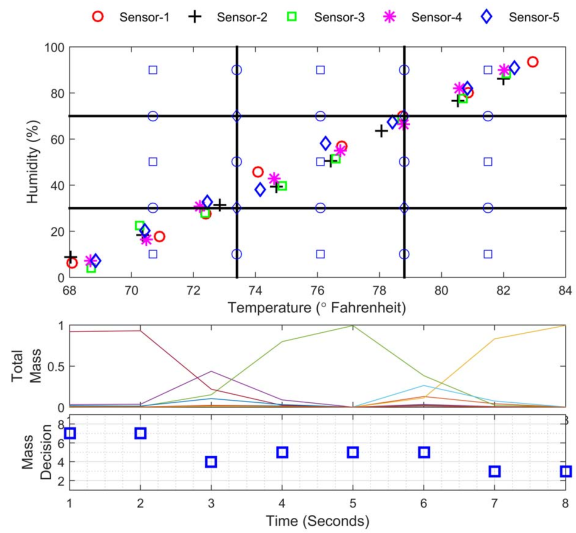

Figure 6. Mass-decision for diagonal test data set - 9 classes DST model

$$m_k(l_j) = \frac{1/d_k^{l_j}}{\sum_{j=1}^{m}(1/d_k^{l_j})} \tag{23}$$

Finally mass functions for each sensor node k related to all n classes are:

$$m_k = \{m_k(l_1), m_k(l_2), ..., m_k(l_n)\} \tag{24}$$

To evaluate our proposed method, we generate several random test data sets and feed them as input to our MATLAB© program to calculate the Dempster-Shafer combination based on Fig.3. One of the test data set is shown in Fig.6. In Fig.6 eight set of random data are chosen that move diagonally from bottom left to right top along the time. First two sets are inside zone seven (VII), next set inside zone four (IV), next three sets are inside comfort zone and the last two sets are in zone three (III). The total mass function and related decision based on the maximum values of mass are also shown. Based on the first model, all nine classes are singletone and exclusive, so the mass and belief are equal. Although the value of plausibility is greater than mass with the value of final uncertainty (small value as an offset), the overall decision result are the same for mass, belief, plausibility and pignistic probability, as expected.

Similarly we feed our experimental data to the proposed algorithm in Fig.3 to calculate the Dempster-Shafer combination. Fig.7 shows the total mass and the related decision. It is observed that, when AC is turned on, four times it moves inside the comfort zone (class 5) from class 6. We calculated related conflict during all ten hours, and conflict values are

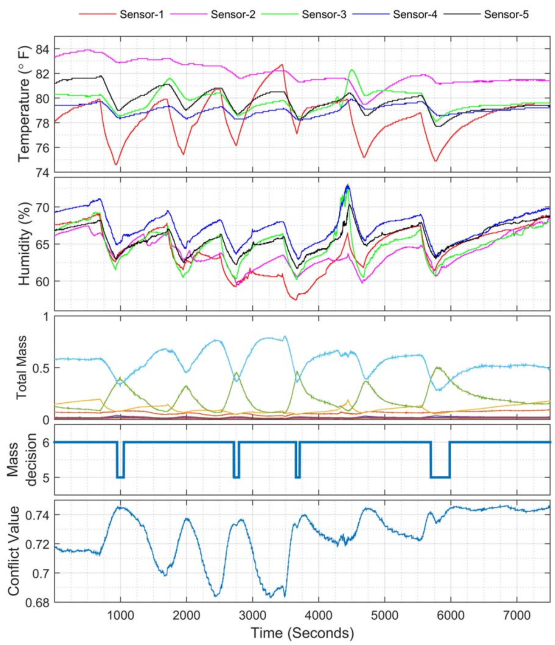

Figure 7. Total mass, decision and conflict for 9 classes DST

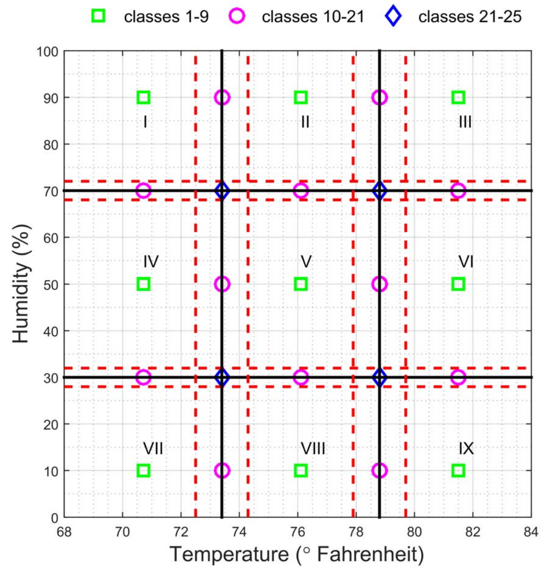

Figure 8. Proposed 25 classes for summer season - Second model

high. Thus it is clear that to overcome the effect of high conflict we need to apply DSmT.

According to the Sensordrone specification document [11], accuracy of temperature sensors are $+/-0.5°C$ or $+/-0.9°F$ and accuracy of humidity sensors are +/- 2%RH. Therefore measurements reported by Sensordrone sensors add uncertainty factor based on the accuracy range of related sensors. Thus we expand our first model to a more accurate one as in Fig.8. Dashed lines in Fig.8 are drawn around solid line, intersection between classes, based on $+/-0.9°F$ and +/- 2%RH measurement error for temperature and humidity sensors, respectively. That means each class can be extended from its solid line boundary to near dashed line. We call this proposed model the second model. We can treat this new proposed model in two different ways, refined Dempster-Shafer model or hybrid DSm model. Because original nine classes are not exclusive completely in the second model, it is not a Dempster-Shafer model any more. In fact this second model is a hybrid DSm model, not a free DSm model, because there are some exclusivity among some classes but not full non-exclusivity among all classes. For example, based on Fig.8 class one has intersection with classes 2,4 and 5 while it is exclusive from classes 3,6,7,8 and 9. It is clear that class 5 is the only class that has intersection with all other classes.

If we define each new decomposed area in Fig.8 as a new class, total 25 classes without any intersection with others, then we can have Dempster-Shafer model with 25 exclusive and exhaustive classes. In this case, Dempster-Shafer combination rule can be applied to calculate total mass functions on the new 25 classes instead of 9 classes before. Note that we only use this as a ground truth in this study. The number of decomposed area without any intersection with each other will grow exponentially when uncertainties exist, thus in reality it would prevent the use of DST with exclusive and exhaustive classes due to the huge number of the decomposed area. On the contrary, the number of the areas remains the same for DSm hybrid model, as explained later.

Fig.9 shows the results for random test data set. It is clear that decision based on belief, plausibilty and pignistic probability are similar and outperform the decision based on the mass function. Fig.10 shows the total mass, belief functions and related decision making result for our ten hours data. According to Fig.10, maximum mass functions move between classes 18 (Intersection between class 3 and 6 $< 36 >$, based on Smarandach codification [12]) and 23 (Intersection among classes 2,3,5 and 6 $< 2356 >$, [12]). Even if we consider just maximum mass among original focal classes one to nine, it is clear that maximum mass move four times between class 6 and comfort zone. As a result, decision making by total mass functions do not give reasonable result. It seems belief, plausibility and pignistic probability functions are better for decision making. It is observed in Fig.10 that when AC turned on six times, decision result based on belief (Similar with plausibility and pignistic probability decision) six times transfer to comfort zone (class 5) from class 6. Hence the decision making by belief, plausibility and pignistic probability in this proposed

Table II
PROPOSED 25 CLASSES FOR SUMMER SEASON

Class No.	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Intersection Classes	1,2	2,3	4,5	5,6	7,8	8,9	1,4	2,5	3,6	4,7	5,8	6,9	1,2,4,5	2,3,5,6	4,5,7,8	5,6,8,9
Temperature ($\circ F$)	73.4	78.8	73.4	78.8	73.4	78.8	70.7	76.1	81.5	70.7	76.1	81.5	73.4	78.8	73.4	78.8
Humidity (%)	90	90	50	50	10	10	70	70	70	30	30	30	70	70	30	30

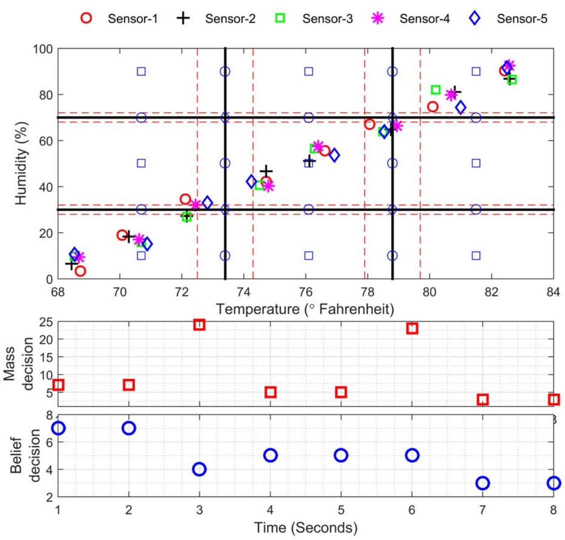

Figure 9. Mass-belief decision for test data set - 25 classes DST model

Table III
AVERAGE RUN-TIME

	DST9	DST25	DSmT
Average Run-Time (Seconds)	0.1601211	0.8071339	4.2721231

second model outperform the first model. Nevertheless, Fig.10 shows that conflict values did not decrease for the new model in DST mode and conflict values are even higher.

As an alternative method, we treat Fig.8 as a DSm hybrid model with nine classes. They are not completely exclusive among all classes but they are exhaustive. We applied PCR5 rule based on the quasi method outlined in Section III. The result in Fig.11 and Fig.12 show that PCR5 decision is very accurate even if there are only nine classes in the DSm hybrid model.

Table.III compares the average run time for the three methods discussed. It is clear that DSmT model with PCR5 needs more computation time in this test case. However, DSmT model will sustain because the number of classes remains the same while DST model will not due to the exponential growth in the number of classes, as explained before. Thus it is expected that DSmT model with PCR5 would be appropriate for big data processing with large number of classes or high cardinality. Furthermore, DSmT model with PCR5 outperforms DST model with the same number of classes by a big margin. Define P_D as the

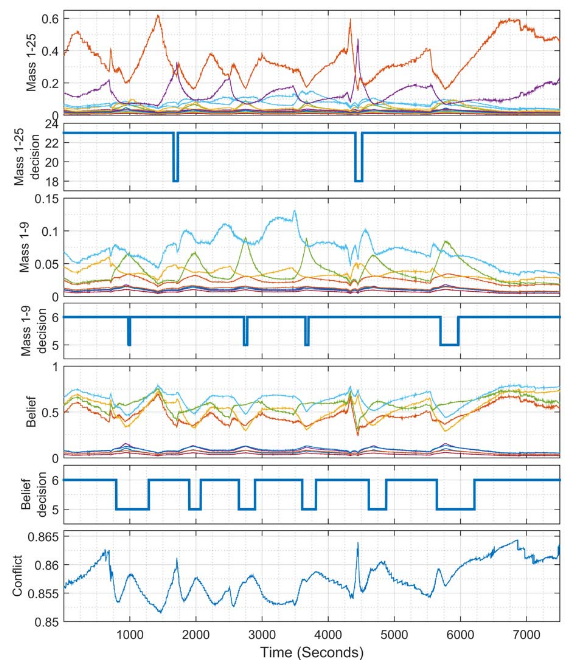

Figure 10. Decision making based on mass and belief for 25 classes DST

Table IV
PROBABILITY OF DETECTION AND FALSE ALARM

Model	P_D	P_F
DST9	26.58%	0
DSmT-PCR5	96.31%	17.38%

detection probability, i.e., correct detection of comfort zone ($P_F = Pr(H_1|H_1)$). Similarly P_F is define as the probability of wrong decision ($P_F = Pr(H_1|H_0)$). Using DST 25 classes model as a ground truth, we calculate P_D and P_F for DST 9 classes and DSmT model. Table.IV shows that DSmT model has much higher $P_D = 96.31\%$ comparing to DST 9 model $P_D = 26.58\%$.

VI. CONCLUSIONS

The feasibility of using Dezert-Smarandache Theory (DSmT) for big data processing is explored in this paper. The methods in DSmT such as PCR5 have very high computational complexity, thus they cannot be directly applied to multiple big data sources with high cardinality. We propose a DSm hybrid model with reduced number of classes and thus low computational cost and evaluate its performance through a case study. Specifically, the proposed method is

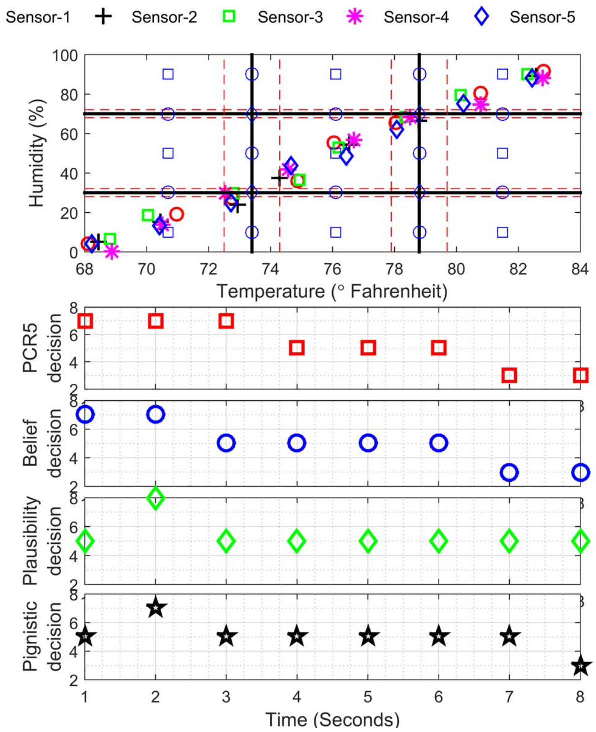

Figure 11. PCR5 decision based on test data for 9 classes DSm model

applied to analyze temperature and humidity data for smart building applications. Our results show that the proposed DSm hybrid model will sustain because the number of classes remains low while DST model will not due to the exponential growth in the number of classes. Comparing to DST with the same number of classes, DSmT has much better performance when the data contain high level of uncertainty. The results using real data sets demonstrate the potentials of the proposed method for big data processing when the data sets contain high level of uncertainty.

VII. Acknowledgments

This research work is supported by the U.S. Office of the Assistant Secretary of Defense for Research and Engineering (OASD(R&E)) under agreement number FA8750-15-2-0119. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Office of the Assistant Secretary of Defense for Research and Engineering (OASD(R&E)) or the U.S. Government.

References

[1] G. A. Shafer, C Mathematical Theory of Evidence," *Princeton University Press*, 1976.

[2] F. Smarandache and J. Dezert, "Advances and Applications of DSmT for Information Fusion," in *American Research Press, Rehoboth*, vol. 2, ch. 1, 2006, pp. 3–68.

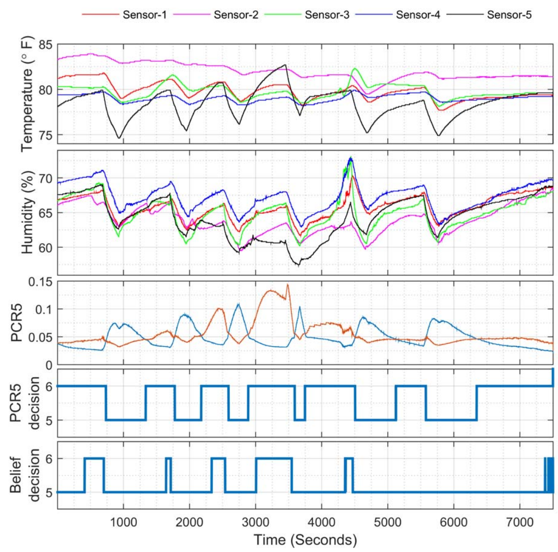

Figure 12. PCR5 and belief decision for 9 classes DSm model

[3] "https://www.ashrae.org/."

[4] H. Jafari, X. Li, L. Qian, and Y. Chen, "Community based sensing: A test bed for environment air quality monitoring using smartphone paired sensors," in *36th IEEE Sarnoff Symposium*, Sep 2015, pp. 12–17.

[5] "http://sensordrone.com/."

[6] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," in *Annals of Mathematical Statistics* vol. 38, 1967, pp.325–339.

[7] R. R. Yager and L. Liu, "Classic Works of the Dempster-Shafer Theory of Belief Functions," *American Research Press, Rehoboth*, vol. 2, ch. 1, 2006, pp. 3–68.

[8] P. Smets, "Constructing the pignistic probability function in a context of uncertainty", *Uncertainty in Artificial Intelligence*, Vol. 5, pp. 29–39, Aug 2004.

[9] F. Smarandache and J. Dezert, "Advances and Applications of DSmT for Information Fusion," in *American Research Press, Rehoboth*, vol. 2, ch. 2, 2006, pp. 69–88.

[10] F. Smarandache and J. Dezert, "An Algorithm for Quasi-Associative and Quasi-Markovian Rules of Combination in Information Fusion," in *5th International Symposium on Applied Computational Intelligence and Informatics*, May 2004, pp. 557–562.

[11] "http://canselsoftware.com/wp-content/uploads/2014/02/"

[12] F. Smarandache and J. Dezert , "Advances and Applications of DSmT for Information Fusion", Vol. 1, ch. 2, 2004, pp. 37–48