



# Inferential Modelling and Decision Making with Data

**DOI:**

[10.23919/IConAC.2017.8082048](https://doi.org/10.23919/IConAC.2017.8082048)

**Document Version**

Final published version

[Link to publication record in Manchester Research Explorer](#)

**Citation for published version (APA):**

Yang, J.-B., & Xu, D. (2017). Inferential Modelling and Decision Making with Data. In *Automation and Computing (ICAC), 2017 23rd International Conference on IEEE*. <https://doi.org/10.23919/IConAC.2017.8082048>

**Published in:**

Automation and Computing (ICAC), 2017 23rd International Conference on

**Citing this paper**

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

**General rights**

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Takedown policy**

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



# Inferential Modelling and Decision Making with Data

Jian-Bo Yang

Decision and Cognitive Sciences Research Centre  
The University of Manchester, Manchester M13 9SS, UK  
Key Laboratory of Process Optimization and Intelligent  
Decision-making, Ministry of Education, Hefei, China  
E-mail: [jian-bo.yang@manchester.ac.uk](mailto:jian-bo.yang@manchester.ac.uk)

Dong-Ling Xu

Decision and Cognitive Sciences Research Centre  
The University of Manchester, Manchester M13 9SS, UK  
Key Laboratory of Process Optimization and Intelligent  
Decision-making, Ministry of Education, Hefei, China  
E-mail: [l.xu@manchester.ac.uk](mailto:l.xu@manchester.ac.uk)

**Abstract** — In this paper, we introduce the main concepts of a new *maximum likelihood evidential reasoning (MAKER)* framework for data-driven inferential modelling and decision making under different types of uncertainty. It consists of two types of model: *state space model (SSM)* and *evidence space model (ESM)*, driven by the data that reflects the relationships between system inputs and output. *SSM* is constructed to describe different system states and changes. *ESM* is established by mapping data to a set of evidence that is partitioned into evidential elements each pointing to a system state set and together represents system behaviours in a probabilistic and distributed manner. The reliability of evidence and interdependence between a pair of evidence are explicitly measured. It is in the joint evidence-state space that multiple pieces of evidence with different degrees of interdependence and reliability are acquired from system inputs and combined to inference system output. A general optimal learning model is constructed, where evidence reliability can be learnt from historical data by maximising the likelihood of true state. In the *MAKER* framework, different types of uncertainty can be taken into account for inferential modelling, probabilistic prediction and decision making.

**Keywords** — *Inferential modelling, Decision making, Evidential reasoning, Likelihood inference, System analysis*

## I. INTRODUCTION

Data generated from routine activities are hardly perfect and are almost always associated with uncertainty (Shafer 1976; Yang and Xu, 2013). It is essential to deal with different types of uncertainty consistently in system modelling and decision making. Conventional data-driven modelling tools such as multivariate regression, support vector machine and artificial neural networks model relationships among system inputs and outputs in a non-probabilistic manner, thus with limited power in handling various types of uncertainty in data, such as randomness, inaccuracy and ambiguity. Logistic regression and Bayesian networks are classical probabilistic modelling tools but are not well suited to handling inaccuracy and ambiguity, which can be caused by factors such as errors and missing values in routinely generated data.

The belief rule based inference methodology based on evidential reasoning (*RIMER*) provides an alternative means

for probabilistic modelling. It uses belief distribution to model random relationships between system inputs and output, attribute and rule weights to take into account data quality, and probability assignment to subsets of system states to model ambiguity. In a *RIMER* model, however, the number of rules increases exponentially with the number of input variables and the referential points of each variable. As such, the number of parameters for training increases exponentially. The evidential reasoning rule (Yang and Xu 2013) can be used to support inferential modelling with imperfect data, but it requires that evidence acquired from data be independent of each other.

This paper aims to introduce a new inference framework, referred to as *maximum likelihood evidential reasoning (MAKER)* framework established recently by the authors for modelling and analysis of complex systems. The *MAKER* framework consists of two types of model: state space model (*SSM*) and evidence space model (*ESM*). *SSM* is constructed to describe system states and changes at different inputs. *ESM* is established by means of a set of interdependent evidence that can represent system behaviours in a probabilistic and distributed manner. In *ESM*, each piece of evidence is acquired from likelihoods that are generated from data, partitioned into evidential elements each associated with a subset of system states, and thus profiled as a basic probability distribution on the power set of system states. The degree of interdependence among a pair of evidence can be measured by marginal and joint likelihood functions. The partition and association of evidence leads to the creation of a joint evidence-state space, in which all evidential elements pointing to the same subset of system states are combined to generate the total support for the state subset. Evidence reliability is proposed to take account of the limited power of an evidential element to assert whether a system state is true or not. In the joint evidence-state space, evidence reliability is defined as the conditional probability to which a system state set is true given that the evidence points to the set, and can be estimated by examining the process where data is generated or by optimal learning from data to maximise the likelihood of true state. In this paper, an optimal learning model is built to estimate evidence reliability from data. After the *SSM* and *ESM* models are optimally trained using historical data, they can be used to infer system output for given inputs by combining multiple pieces of evidence activated from the evidence set of *ESM* by the inputs.

The above-described *MAKER* framework provides a new inferential modelling process for system analysis and decision

making and possesses some unique features that separate it from other modelling frameworks. First, this new inference framework does not depend on prior distribution but admits unknown prior by default. If a prior distribution is available, it is treated as an independent piece of evidence and added to the evidence set of *ESM*. This makes it possible to make use of incomplete data collected from different sources without data imputation or deletion. Secondly, in this new framework, data quality is taken into account by means of evidence reliability. This explicit measurement of data quality not only captures the inaccurate nature of routinely generated data but also leads to the findings of how to conjunctively combine multiple pieces of disjoint or completely conflicting evidence. The third feature is that the interdependence between a pair of evidence is measured by marginal and joint likelihood functions, which can capture the interrelationship between the pair of evidence statistically. Last but not least, representing evidence as basic probability distribution on the power set of system states allows the explicit measurement of ambiguity in data by allowing probabilities to be assigned to subsets of system states, with the recognition that ambiguity can be reduced in the evidence combination process with accumulation of evidence. In the rest of this paper, the new *MAKER* framework is introduced and illustrated using a numerical example.

## II. INTRODUCTION TO *MAKER* FRAMEWORK

From a system point of view, inference is a systematic process of predicting the outputs of a system from its various inputs, in which the correlation, dependence, and interaction among the inputs and between the inputs and outputs of the system need to be taken into account. If an explicit functional relationship between inputs and outputs is known *a priori*, inference could be as direct as to use the function to calculate outputs for given inputs; otherwise, data or judgements about system behaviours need to be gathered to enable inference. The *MAKER* framework is a data-driven inference process from input variables to output variable under uncertainty. In this section, the *MAKER* framework is introduced for data-driven modelling and inference. It consists of two types of model: state space model and evidence space model, and a conjunctive evidential reasoning rule.

### A. State Space Model

In a state space model (*SSM*), an output variable is characterised by a number of states, forming a system state space. The mechanism of *SSM* is most easily introduced by assuming a state space that consists of a finite number of states. Dempster's original thinking on state space (Dempster, 2008) lays the foundation of *SSM* as discussed in this subsection.

Let  $H_n$  stand for a system state. Without loss of generality, suppose a system state space has at least  $N$  disjoint states that do not overlap each other, denoted by  $\Theta = \{H_1 \cdots H_n \cdots H_N \cdots\}$  with  $H_i \cap H_j = \emptyset$  for any  $i \neq j$ . In this paper, probability is allowed to be assigned to a subset of system states. The collection of all subsets is referred to as the power set of  $\Theta$ , denoted as  $P(\Theta)$  or  $2^\Theta$ . The full power set of  $\Theta$  includes the empty set  $\emptyset$  and the full state space  $\Theta$ .

An output is modelled by a unique set function, called basic probability function. A basic probability function is defined as an ordinary discrete probability distribution over the nonempty subsets of  $\Theta$  as follows.

**Definition 1 (Basic probability function).** A set function  $p: 2^\Theta \rightarrow [0, 1]$  is called a basic probability function if the following conditions are met

$$0 \leq p(\theta) \leq 1 \quad \forall \theta \subseteq \Theta \quad (1)$$

$$\sum_{\theta \subseteq \Theta} p(\theta) = 1 \quad (2)$$

$$p(\emptyset) = 0 \quad (3)$$

where  $\theta$  is a subset of states, referred to as an assertion, and  $p(\theta)$  is the basic probability that assertion  $\theta$  is true.  $p(\theta)$  is an atom in that it is committed exactly to assertion  $\theta$  and cannot be broken down into pieces assigned to subsets of  $\theta$ .

**Definition 2 (System output).** A system output  $y$  is profiled as a probability distribution as follows

$$y = \left\{ (\theta, p(\theta)), \forall \theta \subseteq \Theta \text{ and } \sum_{\theta \subseteq \Theta} p(\theta) = 1 \right\} \quad (4)$$

where  $p(\theta)$  is generated from system inputs using Equations (14) and (15).  $\theta$  is called the focal element of  $y$  if  $p(\theta) > 0$ .

Although a basic probability distribution provides a unique and panoramic description for an output as a whole, each assertion in an output carry more detailed information for further analysis. According to Dempster (2008), let  $p'(\theta)$  be probability that quantifies the uncertainty about assertion  $\theta$ , representing the evidence for the truth of  $\theta$ .  $p'(\theta)$  is given by

$$p'(\theta) = \sum_{A \subseteq \theta} p(A) \quad (5)$$

That is,  $p'(\theta)$  accumulates the basic probabilities for all the subsets of states that belong to  $\theta$  and represents the probability that is unambiguously assigned to  $\theta$ . As such,  $p'(\theta)$  is the successor of ordinary probability to which assertion  $\theta$  is true.

On the other hand, let  $p^f(\theta)$  be the probability that measures the evidence against the truth of  $\theta$ , or the degree to which  $\theta$  is believed to be false.  $p^f(\theta)$  can be calculated by  $p^f(\theta) = p'(\theta^c) = \sum_{B \cap \theta = \emptyset} p(B)$  which is the probability for the truth of the negation of  $\theta$ , denoted by  $\theta^c$ .  $p'(\theta)$  and  $p^f(\theta)$  given above are so defined that the assignment of nonzero probability is allowed to ambiguity or "unknown", denoted by  $p^u$ . In summary, an assertion can be described by three nonnegative probabilities ( $p'$ ,  $p^f$ ,  $p^u$ ) with  $p' + p^f + p^u = 1$  assigned to the triad of an assertion being "true", "false" and "unknown". In other words, *MAKER* is not restricted to  $p'$  and  $p^f$  with  $p' + p^f = 1$  as in Bayesian theory, but allows  $p^u = 1 - p' - p^f$  to quantify the ambiguity so that inference can be conducted with ambiguous information. In *SSM*, while an output is profiled by a unique

basic probability function, every member of the power set in the output is described in more detail by a  $(p^t, p^f, p^u)$  triplet.

### B. Evidence Space Model

Whether an assertion in the state space is true or not depends on whether there is evidence supporting this assertion or not. An evidence space is a derived space from data in the sense that each piece of evidence in the evidence space is acquired from data. In the context of using evidence to support assertions, each piece of evidence is partitioned into evidential elements, each of which points exactly to an assertion in the state space (or an element in the power set of the states). Let  $e_{i,l}$  be the  $i^{\text{th}}$  piece of evidence from the  $l^{\text{th}}$  input variable  $x_l$  and  $e_{i,l}(\theta)$  an element of  $e_{i,l}$ , pointing exactly to assertion  $\theta$ .

A process for acquiring evidence from data for a single input variable was established on the basis of the likelihood principle and the Bayesian principle that the combination of  $e_{i,l}$  with the prior distribution of states should lead to posterior probability (Yang and Xu, 2014). Based on this evidence acquisition process, a one-dimensional *ESM* can be built for each input variable; the basic probability that evidence  $e_{i,l}$  points to assertion  $\theta$  is given by  $p_{\theta,i,l} = p_l(e_{i,l}(\theta))$ . Suppose  $c_{\theta,i,l}$  is the likelihood that the  $i^{\text{th}}$  value of  $x_l$  is observed given assertion  $\theta$ . To enable probabilistic inference,  $p_{\theta,i,l}$  ought to be acquired from normalized likelihood as follows:

$$p_{\theta,i,l} = c_{\theta,i,l} / \sum_{A \subseteq \Theta} c_{A,i,l} \quad \forall \theta \subseteq \Theta \quad (6)$$

Given that basic probability  $p_{\theta,i,l}$  is acquired from input  $x_l$  for each  $\theta$ , evidence  $e_{i,l}$  can be used as system input and represented as the following probability distribution :

**Definition 3 (System input).** A system input is profiled as a basic probability distribution as follows

$$e_{i,l} = \left\{ (e_{i,l}(\theta), p_{\theta,i,l}), \forall \theta \subseteq \Theta \text{ and } \sum_{\theta \subseteq \Theta} p_{\theta,i,l} = 1 \right\} \quad (7)$$

where  $p_{\theta,i,l}$  is generated from input variable  $x_l$  by Equation (6) and evidential elements  $e_{i,l}(H_n)$  for all  $H_n \in \Theta$  constitute the evidence subspace for the  $i^{\text{th}}$  value of  $x_l$ .

If  $x_l$  is discrete, a piece of evidence can be acquired for each of its values, leading to a discrete *ESM* for  $x_l$ , denoted by evidence subset  $E_i = \{e_{1,l}, e_{2,l}, \dots, e_{i,l}, \dots\}$ . The *ESM* represents the probabilistic relationship between  $x_l$  and output  $y$  in a distributed manner. If  $x_l$  is continuous, there can be different ways to build a *ESM* for  $x_l$ . An obvious one is to discretise  $x_l$  and follow the above procedure. Alternatively, a continuous likelihood density function can be estimated and used to generate a basic probability distribution for each value of  $x_l$ , leading to a continuous *ESM* for  $x_l$ . Given the limited space of the short paper, the continuous case is not elaborated here.

In the evidence space, a key question is how to characterize the interrelationship between each pair of input variables when there are multiple inputs. In particular, we are interested in the statistical interdependence between a pair of evidence acquired from two different input variables. An interdependence index is defined in this subsection. Based on the likelihood principle and the Bayesian principle, it can be shown that joint basic probability can be acquired from joint likelihood function. Let  $e_{i,l}$  and  $e_{j,m}$  be two pieces of evidence acquired from two input variables  $x_l$  and  $x_m$  at  $x_l = x_{i,l}$  and  $x_m = x_{j,m}$  respectively. If  $c_{\theta,il,jm}$  is the joint likelihood that both  $x_{i,l}$  and  $x_{j,m}$  are observed given assertion  $\theta$ , then the joint basic probability that both  $x_l = x_{i,l}$  and  $x_m = x_{j,m}$  point to  $\theta$  is given by

$$p_{\theta,il,jm} = c_{\theta,il,jm} / \sum_{A \subseteq \Theta} c_{A,il,jm} \quad \forall \theta \subseteq \Theta \quad (8)$$

Given the above discussions, we are now in a position to analyse the interrelationship between two pieces of evidence by defining an interdependence index as follows. The degree of interdependence between evidential elements  $e_{i,l}(A)$  and  $e_{j,m}(B)$  is defined by  $\alpha_{A,B,i,j}$  as follows:

$$\alpha_{A,B,i,j} = \begin{cases} 0 & \text{if } p_{A,i,l} = 0 \text{ or } p_{B,j,m} = 0 \\ p_{A,B,il,jm} / (p_{A,i,l} p_{B,j,m}) & \text{otherwise} \end{cases} \quad (9)$$

Interdependence index  $\alpha_{A,B,i,j}$  has the following property

$$\alpha_{A,B,i,j} = \begin{cases} 0 & \text{if } e_{i,l}(A) \text{ and } e_{j,m}(B) \text{ are disjoint} \\ 1 & \text{if } e_{i,l}(A) \text{ and } e_{j,m}(B) \text{ are independent} \end{cases} \quad (10)$$

Note that the above interdependence index is defined in the evidence space where basic probability is acquired as normalized likelihood. To interpret the index in terms of ordinary likelihood, it needs to be scaled by multiplying the following constant that does not change with  $\theta$ , to factor in the effect of the normalization process given in Equation (6) and Equation (8)

$$k_{i,j} = \sum_{\theta \subseteq \Theta} c_{\theta,i,l} \sum_{\theta \subseteq \Theta} c_{\theta,j,m} / \sum_{\theta \subseteq \Theta} c_{\theta,il,jm} \quad (11)$$

### C. Conjunctive MAKER Rule

In the *MAKER* framework, it is in the joint evidence-state space that output is inferred from inputs probabilistically. The inference is conducted at elementary level where each assertion in the state space intersects with every evidential element in the evidence space. Such intersection leads to identification of elementary support for an assertion from evidential elements individually. Joint support for an assertion from a conjunction of evidential elements that depend on each other can also be identified. It is the exhaust combination of all such elementary support for an assertion that leads to the establishment of a conjunctive *MAKER* rule to generate the degree of the total support for the assertion. In this subsection, the *MAKER* rule is introduced and the basic ideas on how the inference is performed are discussed.

By definition,  $e_{i,l}(\theta)$  denotes that the observation of  $x_{i,l}$  or evidence  $e_{i,l}$  points exactly to assertion  $\theta$ . The intersection of  $\theta$  and  $e_{i,l}(\theta)$ , or  $s_{i,l}(\theta) = \theta \cap e_{i,l}(\theta)$ , means that evidence  $e_{i,l}$  points to  $\theta$  and  $\theta$  is true as supported by  $e_{i,l}$ , or  $\theta$  is supported by  $e_{i,l}$  for short. Similarly,  $s_{j,m}(\theta) = \theta \cap e_{j,m}(\theta)$  means that  $\theta$  is supported by evidence  $e_{j,m}$ . A question then arises as to how to combine the two pieces of support to generate the combined support for  $\theta$ . To answer the question, we first discuss how to measure support from evidence that can be acquired from different data sources.

If evidence  $e_{i,l}$  is acquired from a data source common to all evidence, characterized by a probability function  $p$ , then the probability mass that assertion  $\theta$  is supported by  $e_{i,l}(\theta)$  is given as follows

$$p(s_{i,l}(\theta)) = p(\theta|e_{i,l}(\theta))p(e_{i,l}(\theta)) = r_{\theta,i,l}p(e_{i,l}(\theta)) \quad (12)$$

In the above equation,  $r_{\theta,i,l} = p(\theta|e_{i,l}(\theta))$  is referred to as the reliability of evidential element  $e_{i,l}(\theta)$  in the sense that it is defined as the conditional probability to which assertion  $\theta$  is true given that  $e_{i,l}$  points to  $\theta$ .  $r_{\theta,i,l}$  in essence measures the quality of  $e_{i,l}$ , which is concerned with how data is generated and how  $e_{i,l}$  is acquired from the data.

If  $e_{i,l}$  is acquired from a data source that is different from other evidence, characterized by a probability function  $p_l$ , the basic probability that  $e_{i,l}$  points to  $\theta$  is given by  $p_l(e_{i,l}(\theta))$  and the probability mass that evidence  $e_{i,l}$  supports  $\theta$  is given by  $p_l(s_{i,l}(\theta))$  with  $s_{i,l}(\theta) = \theta \cap e_{i,l}(\theta)$ . If  $e_{i,l}$  is measured by a likelihood function as in Equation (6), according to the likelihood principle, its support for  $\theta$  should have the same meaning whether it is measured by  $p$  or  $p_l$ . In other words,  $p(s_{i,l}(\theta))$  and  $p_l(s_{i,l}(\theta))$  should be proportional to each other, or  $p(s_{i,l}(\theta)) = \omega_{i,l}p_l(s_{i,l}(\theta))$ , where  $\omega_{i,l}$  is a positive scaling factor constant for all  $\theta \in \Theta$ . Let

$$\begin{aligned} m_{\theta,i,l} &= p(s_{i,l}(\theta)) = \omega_{i,l}p_l(s_{i,l}(\theta)) \\ &= \omega_{i,l}p_l(\theta \cap e_{i,l}(\theta)) = \omega_{i,l}p_l(\theta|e_{i,l}(\theta))p_l(e_{i,l}(\theta)) \quad (13) \\ &= w_{\theta,i,l}p_l(e_{i,l}(\theta)) \end{aligned}$$

In the above equation,  $w_{\theta,i,l} = \omega_{i,l}p_l(\theta|e_{i,l}(\theta))$  is referred to as the weight of evidential element  $e_{i,l}(\theta)$  in the sense that it is given to be proportional to the conditional probability to which  $\theta$  is true given that  $e_{i,l}$  points to  $\theta$ , which is measured by a probability function  $p_l$  constructed from data for  $x_l$  only. Note that  $w_{\theta,i,l} = r_{\theta,i,l}$  when  $p = p_l$ , which means  $\omega_{i,l} = 1$ .

Given the above discussions and definitions, we are now in a position to introduce the conjunctive *MAKER* rule for generating the combined probability that assertion  $\theta$  is jointly supported by both pieces of evidence  $e_{i,l}$  and  $e_{j,m}$  as follows

$$p(\theta) = \begin{cases} 0 & \theta = \emptyset \\ m_\theta / \sum_{C \subseteq \Theta} m_C & \theta \subseteq \Theta \end{cases} \quad (14)$$

where  $m_\theta$  measures the combined probability mass for  $\theta$  from both  $e_{i,l}$  and  $e_{j,m}$ , given as follows

$$\begin{aligned} m_\theta &= \left[ (1-r_{j,m})m_{\theta,i,l} + (1-r_{i,l})m_{\theta,j,m} \right] \\ &+ \sum_{A \cap B = \theta} \gamma_{A,B,i,j} \alpha_{A,B,i,j} m_{A,i,l} m_{B,j,m} \end{aligned} \quad (15)$$

where  $r_{i,l} = \sum_{\theta \subseteq \Theta} r_{\theta,i,l}p(e_{i,l}(\theta))$  is the reliability of evidence  $e_{i,l}$ .  $\gamma_{A,B,i,j}$  is a nonnegative coefficient, which reflects the degree of joint support for  $\theta$  from both  $e_{i,l}$  and  $e_{j,m}$  relative to their individual support given that  $e_{i,l}$  points to assertion  $A$  and  $e_{j,m}$  points to assertion  $B$ .  $\gamma_{A,B,i,j}$  is assumed to be 1 in the following numerical example. Equation (15) should be applied for evidence recursively before Equation (14) is implemented.

Note that the above general conjunctive *MAKER* rule can result in specific inference rules under various conditions. In particular, if  $w_{\theta,i,l} = w_{i,l}$  for any  $A, B, \theta \subseteq \Theta$ , the conjunctive *MAKER* rule applied for combination of independent evidence reduces to the evidential reasoning rule (Yang and Xu, 2013). If it is further assumed that  $r_{i,l} = w_{i,l} = 1$ , the rule reduces to Dempster's rule (Shafer 1976), which in turn reduces to Bayes' rule when there is no ambiguity in data.

#### D. General *MAKER* Training Model

In Equation (15),  $r_{\theta,i,l}$ ,  $w_{\theta,i,l}$ ,  $r_{\theta,j,m}$ ,  $w_{\theta,j,m}$  and  $\gamma_{A,B,i,j}$  are the parameters that need to be assigned for inference. Generally, the values of these parameters can be trained using input-output datasets. The trained parameters can then be used to predict system outputs for given inputs.

A general least square optimization model is established as follows, to set the parameters by maximizing the likelihood of true state:

$$\min \delta = \frac{1}{2S} \sum_{s=1}^S \sum_{\theta \subseteq \Theta} (p(\theta) - \hat{p}^{(s)}(\theta))^2 \quad (16)$$

$$s.t. \quad r_{\theta,i,l}, w_{\theta,i,l}, \gamma_{A,B,i,j} \in \Omega$$

In the above optimization model,  $\hat{p}^{(s)}(\theta)$  is the probability that assertion  $\theta$  is true, as given in the  $s^{\text{th}}$  observation; the objective is to make  $p(\theta)$  as close to  $\hat{p}^{(s)}(\theta)$  as possible, or to maximize the likelihood of true state; there is  $0 \leq \delta \leq 1$ .  $\Omega$  is the feasible space of parameters, e.g.  $\varepsilon \leq r_{\theta,i,l} \leq 1$  by definition with  $\varepsilon$  being a nonnegative low bound set according to the accuracy of data and the process of acquiring evidence from data. Other constraints can also be included in  $\Omega$ .

### III. NUMERICAL EXAMPLE

#### A. Problem Description and Data Collection

This illustrative example is about the diagnosis of a disease given results from two types of test: saliva test and blood test. Suppose routinely generated data records are collected, as shown in Table 1 to Table 3 where there are two system inputs and one system output. The output ( $y$ ) is whether it is true ( $y_1$ ), false ( $y_2$ ) or unknown that a patient has the disease; the inputs are saliva test ( $x_1$ ) and blood test ( $x_2$ ); each data record shows whether  $y$  is true, false or unknown given positive ( $x_{1,1}$ ) or negative ( $x_{2,1}$ ) saliva test and positive ( $x_{1,2}$ ) or negative ( $x_{2,2}$ ) blood test. In this example, the records with the same input and output values are grouped together, and the input and output variables each have only two different categorical values.

Table 1 Joint Data Records for Both Tests

Data Record No.	Input variables		Output Disease ( $y$ )
	Saliva ( $x_1$ )	Blood ( $x_2$ )	
1-93	Positive	Positive	True
94-95	Positive	Negative	True
96-99	Negative	Positive	True
100	Negative	Negative	True
101-120	Positive	Positive	False
121-160	Positive	Negative	False
161-180	Negative	Positive	False
181-1000	Negative	Negative	False
1001-1043	Positive	Positive	Unknown
1044-1045	Positive	Negative	Unknown
1046-1048	Negative	Positive	Unknown
1049-1100	Negative	Negative	Unknown

Table 2 Individual Data Records for Saliva Test Only

Data Record No.	Saliva Test ( $x_1$ )	Disease ( $y$ )
1101-1280	Positive	True
1281-1295	Negative	True
1296-1450	Positive	False
1451-3100	Negative	False
3101-3150	Positive	Unknown
3151-3200	Negative	Unknown

Table 3 Individual Data Records for Blood Test Only

Data Record No.	Blood Test ( $x_2$ )	Disease ( $y$ )
3201-3450	Positive	True
3451-3470	Negative	True
3471-3610	Positive	False
3611-6200	Negative	False
6201-6254	Positive	Unknown
6255-6300	Negative	Unknown

In Table 1, the first 1000 records are complete in the sense that they each record the results of saliva test, blood test and disease diagnosis; the last 100 records are ambiguous due to the unknown result of disease diagnosis. The data records in Table 2 and Table 3 are incomplete and ambiguous as only

saliva test or blood test is recorded and some of the diagnosis results for the disease are unknown.

A conventional method is to use only complete and unambiguous data for modelling and discard ambiguous and incomplete data, or use data imputation to estimate missing values. However, the former approach leads to the loss of information and probably leads to biased or misleading results; the latter approach creates non-existing data and as thus is prone to errors. In this section, we show how to use ambiguous and incomplete data together with complete data for modelling, prediction and decision making.

#### B. Acquisition of Evidence from Data Records

The data records shown in Table 1 to Table 3 can be used to acquire evidence when the result of saliva test or blood test is positive or negative. The first step of evidence acquisition is to generate a frequency table for each input variable from the data records, for example Table 4 for saliva test where both joint and individual data records are used for modelling. The likelihoods that saliva test is positive or negative given that the disease is true, false or unknown are calculated from Tables 4, as shown in the  $c_{\theta,1,1}$  and  $c_{\theta,2,1}$  columns of Table 5. The basic probability that a positive or negative saliva test points to the true, false or unknown disease can then be generated by using Equation (6), leading to the construction of a subset of evidence for saliva test. In this example, this subset of evidence has two pieces of evidence as shown in the  $p_{\theta,1,1}$  and  $p_{\theta,2,1}$  columns of Table 5. For example, if there is a result that saliva test for a patient is positive, this evidence points to the patient having the disease with a large probability of 0.627 and not having it with a very small probability of 0.0535, whilst there exists a significant degree of ambiguity of 0.3195. Similarly, a subset of evidence for blood test can be established.

Table 4 Frequency using all data for saliva test ( $x_1$ )

Frequency		Positive ( $x_{1,1}$ )	Negative ( $x_{2,1}$ )
Disease	True ( $y_1$ )	275	20
	False ( $y_2$ )	215	2490
	Unknown ( $\theta$ )	95	105

Table 5 Likelihood and basic probability for saliva test

Likelihood   basic probability		Positive ( $x_{1,1}$ )		Negative ( $x_{2,1}$ )	
		$c_{\theta,1,1}$	$p_{\theta,1,1}$	$c_{\theta,2,1}$	$p_{\theta,2,1}$
Disease	$e_1(\theta = y_1)$	0.9322	0.6270	0.0678	0.0448
	$e_1(\theta = y_2)$	0.0795	0.0535	0.9205	0.6083
	$e_1(\theta = \theta)$	0.4750	0.3195	0.5250	0.3469

#### C. Analysis of Evidence Interdependence

If a patient has both saliva test and blood tests, there is a need to combine the two pieces of evidence to provide a more accurate diagnosis than what an individual test result can tell. A conventional approach is to use complete joint data records, such as the first 1000 data records in Table 1, to construct a

joint probability table to find the probabilities of the patient having disease or not, and use the probabilities to advise the patient accordingly. However, questions can be raised as to whether such advice is reliable or not. For example, is it true that the prior distribution of the patient having or not having the disease is 0.1 or 0.9 respectively as shown by the complete data in Table 1? If not, such advice can lead to underestimation or overestimation of the probabilities. If a large number of data records for individual saliva or blood test are not used, would this affect the accuracy or reliability of the probabilities? Also, how does the omission of the ambiguous data affect the probabilities? To address these concerns, we illustrate the next steps of using the *MAKER* framework to advise the patient.

In the *MAKER* framework, joint data records are used to estimate interrelationship between each pair of evidence. To do so, joint basic probability is estimated from Table 1 using Equation (8), as shown in Table 6. For example, for a patient who has positive saliva test result ( $x_{1,1}$ ) and positive blood test result ( $x_{1,2}$ ), based on the data records of Table 1, his chance of having the disease or not may be estimated to be 0.6728 and 0.0161 respectively, though there exists a degree of ambiguity of 0.3111, independent of the prior that may be estimated from Table 1. In the *MAKER* framework, the joint basic probability is only used to generate interdependence indices between each pair of evidential elements, as shown in Table 7 which tells that saliva test and blood test are highly interdependent for diagnosis of the disease, as there is always  $\alpha_{\theta,i,j} > 1$  whether either of the test results is positive or not. For example, the two tests are moderately interdependent upon each other for diagnosis of the disease being true when both results are positive as  $\alpha_{y_1,1,1} = 1.579$  ( $x_{1,1}$  and  $x_{1,2}$ ); however, they become very highly interdependent upon each other for the diagnosis when saliva test is negative ( $x_{2,1}$ ) and blood test is positive ( $x_{1,2}$ ) as  $\alpha_{y_1,2,1} = 20.218$  in this case.

Table 6 Estimate joint basic probability from Table 1

Estimates of joint basic probability		$x_{1,1}$		$x_{2,1}$	
		$x_{1,2}$	$x_{2,2}$	$x_{1,2}$	$x_{2,2}$
Disease	$e_{1,2}(y_1)$	0.6728	0.2368	0.4337	0.0069
	$e_{1,2}(y_2)$	0.0161	0.5263	0.2410	0.6322
	$e_{1,2}(\theta)$	0.3111	0.2368	0.3253	0.3608

Table 7 Interrelationship measurement between the tests

Interdependence index $\alpha_{\theta,i,j}$		$x_{1,1}$		$x_{2,1}$	
		$x_{1,2}$	$x_{2,2}$	$x_{1,2}$	$x_{2,2}$
Disease	$e_{1,2}(y_1)$	1.579	18.594	20.218	10.821
	$e_{1,2}(y_2)$	11.734	18.486	13.133	1.658
	$e_{1,2}(\theta)$	3.250	2.181	2.907	2.842

#### D. Maximum Likelihood Inference and Decision Making

To provide realistic diagnosis from saliva test and blood test, individual evidence with reliability can be combined using the *MAKER* rule of Equation (15). If the reliabilities of disease diagnoses from saliva and blood tests and the prior distribution are estimated using the model given in Equation (16) with all the 6300 data records and  $\varepsilon = 0.5$ , the maximum likelihood reliabilities are generated as in Table 8 and the combined joint basic probabilities, independent of the prior, are given in Table 9 with  $\delta = 0.0784$ . Compared with the estimates of joint basic probabilities as shown in Table 6, the results in Table 9 have smaller combined joint probabilities for the unknown state. This is due to two reasons. First, the combination of unknown diagnosis from one test with true or false diagnosis from another test leads to the reduction of probability mass for the unknown state in Equation (15). Secondly, the reliabilities optimised for unknown diagnosis are low, which reduces the contribution of evidential elements for the unknown state further during the combination process. This provides a solid basis for robust decision making based on both test results.

Table 8 Maximum likelihood reliability of evidence

Reliability	Saliva		Blood		Prior
	Positive	Negative	Positive	Negative	
True	1.0000	0.5000	1.0000	0.5000	0.5000
False	0.9611	1.0000	0.5000	1.0000	0.7405
Unknown	0.8058	0.5000	0.6663	0.5000	0.5000

Table 9 Optimal combined joint basic probability

Joint basic probability		$x_{1,1}$		$x_{2,1}$	
		$x_{1,2}$	$x_{2,2}$	$x_{1,2}$	$x_{2,2}$
Disease	$y_1$	0.8218	0.2916	0.4861	0.0090
	$y_2$	0.0201	0.5839	0.3372	0.8651
	$\theta$	0.1581	0.1245	0.1767	0.1258

#### E. Concluding Remarks

In this short paper, the *MAKER* framework is briefly introduced with an illustrative example. It is a new data-driven process for modelling complex system behaviors under various types of uncertainty, based on the establishment of an evidence space with the interrelationship between each pair of evidence explicitly measured and the reliability of evidence optimally estimated under the principle of maximum likelihood.

#### REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.
- [2] A. P. Dempster, "The Dempster-Shafer calculus for statisticians". *International Journal of Approximate Reasoning*, Vol.48, 2008, pp.365-377.
- [3] J. B. Yang and D. L. Xu, "Evidential reasoning rule for evidence combination", *Artificial Intelligence*, Vol.205, 2013, pp.1-29.
- [4] J. B. Yang and D. L. Xu, "A study on generalising Bayesian inference to evidential reasoning". In *Belief Function: Theory and Applications*, Lecture Notes in Artificial Intelligence, Fabio Cuzzolin (ed.), 2014.