IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 13, Issue 5 Ver. III (Sep. - Oct. 2017), PP 32-36
www.iosrjournals.org

On a Novel Method to Measure Semantic Information through Possibilistic Restrictions

Jayesh Karanjgaonkar [1]& Purushottam Jha [2]
[1] Govt. Sukhram Nage College, Nagri, C.G., India
[2] Govt. Dau Kalyan P.G. College, Baloda-Bazar, C.G., India
Corresponding Author: Jayesh Karanjgaonkar

***Abstract**: We present a novel method to fit a possibility distribution on a continuous stream data system and hence fitting a restriction on the same by using mode of the data. This method is an extension of Zadeh's restriction centered theory of information. The proposed method is suitable to measure and analyse information in complex or hybrid systems, where data sufficiency is never fulfilled and approximate reasoning is required. This method mimics the perception calculation technique, which is deeply ingrained in human mind, by which it gathers and analyse information in natural language. A method to precisiate the restriction on the basis of difference between probability and possibility summation is also proposed.*

***Keywords**- Possibility Distribution, possibilistic restriction, Semantic Information*
***AMS Subject Classification** – 94D05*

Date of Submission: 03-10-2017 Date of acceptance: 18-10-2017

I. Introduction

In 1949 W. Weaver [8] presented his ideas on problems of communication by classifying them in three categories, namely technical, semantic and control. Technical aspects concern with the coding, decoding and compression of symbols, texts, message and hence *information*, while semantic aspects deal with the meaning of the message or information communicated. The third class became a distinct discipline in which various types of control systems (human, ma- chine and hybrid) are studied. The first class is concerned with the sender of the message while second class is concerned with the receiver of the message. All three classifications one way or another covers every subject in which information/communication is important, for example Tele-communication engineering, Computer Networks, Economics, Psychology, Artificial Intelligence, Decision theory etc.

Recent literature pertaining to information theory shows the interest in search of a theory which will enable us to understand information/ message or meaning of the communication [4, 10, 14,15,17,18,19]. The meaning part becomes essential when we consider various human - human or human - machine systems, for example economic activity, which is regulated by financial / economic data and its interpretations by economic agents. There are many theories to define and quantify semantic content of information or propose changes in mathematical information theory to make it more accommodating for meaning part. Measuring information through possibility distribution is a prominent theory for the same.

II. Possibility Distributions

Zadeh defined theory of possibility on the basis of fuzzy sets and related it to information measurement [13]. Successive improvements were made by Klir and Higashi [6], Dubois & Prade [1, 2, 5, 3], Yager [9, 11] and many more. However the term possibility was coined by Shackle in 1950 in Economics as the potential surprise of an event. In information theory the application of possibility theory was first recognized by Dampster and Shafer [7] in their theory of evidence, which uses pair of dual measures of possibility and necessity to define information.

A possibility distribution or measure Π on a set X is defined by a mapping $\pi : P(X) \rightarrow [0, 1]$, where $P(X)$ being power set of X , with the following properties -

1. $\pi(\phi) = 0$
2. $A \leq B \Rightarrow \pi(A) \leq \pi(B)$
3. $\pi(A \cup B) = \max\{\pi(A), \pi(B)\}$, where $A \cap B = \phi$

Following points are important in the context of possibility, for $Ai \subset X$

- *Normalization* - A possibility distribution is said be normal if $\pi(A_i) = 1$ for at least one $Ai \subset X$.
- *Complete Knowledge* – When $\pi(A_i) = 1$ and $\pi(A_j) = 0$ for $i \neq j$

- *Complete Ignorance* - When $\pi(A_i) = 1$, for $\forall\ i$.
- When $0 \leq \pi(A) \leq 1$, for $A \subset X$, then A is possibly uncertain.
- The dual of possibility, known as necessity is defined as $\eta(A) = 1 - \pi(A^c)$, where A^c is complement of $A \subset$ X

A possibility distribution is characterized by a function f : X → [0, 1]

$$\pi(A) = \sup_{x \in A} f(x),\ A \subset X \qquad (1)$$

A more informative possibility distribution follows minimum specificity criterion. [2, 5, 9]

2.1 Relation to Probability

According to Zadeh [13], if $P(A_i)$ is the probability and $\pi(A_i)$ is the possibility of $A_i \subset X$, then $P(A_i) \leq \pi(A_i)$, which is termed as "*what is probable, that is possible*". The difference between a possibility distribution and a probability distribution is that probability distribution satisfies additivity condition $P(A \cup B) = P(A) + P(B)$, while possibility is non-additive and satisfies maximality condition $\pi(A \cup B) = \text{Max}\{\pi(A), \pi(B)\}$. There is also a possibility - probability consistency principle which states that, for a constant C –

$$\sum_{i=1}^{n} P(A_i)\,\pi(A_i) = C \qquad (2)$$

A possibilistic restriction "X is A" represents a possibilistic distribution defined by $\pi(X = u) = \mu A(u)$, where A is a fuzzy set defined by a membership function μ_A in set of discourse U. The information $I(p)$ conveyed by a restriction "X is A" is equivalent to the possibility distribution $\Pi(p) = \pi(X = u) = \mu A(u)$.
Hence we can write $I(p) \equiv \Pi(p)$, thus the possibility distribution is another method to denote and measure information [12, 13].

A more informative possibility distribution follows minimum specificity criterion. [2, 5, 9]

2.2 Information as Restrictions

Recently Zadeh proposed the Information Principle [20], according to which information is communicated as restrictions. A restriction is a generalized constraint which restricts the values of a variable. Let X be a variable and R be a set of values then a restriction is defined by "*XisrR*", where *isr*, read as "*izar*", is the restriction type. All of the entities X, R and r can be n-arry. When r is blank, i.e. the constraint is "X is R" is known as a Possibilistic Restriction. Similarly we can define probabilistic restriction "X *isp* P" or a bimodal restriction, which is combination of probabilistic and possibilistic restriction.

In most cases X and R are defined in natural language and hence the possibilistic restriction becomes fuzzy possibilistic restriction. If the set R is crisp and singular it becomes an equation, however when R is non-singular (and fuzzy) the possibilistic restriction implies uncertainty, hence it can represents information. A more restricted restriction represents more information (minimum specificity). Since the variables are defined in natural language, possibilistic restrictions are best way to capture and analyse semantic information i.e. meaning and information [15-20]. We make above definitions clear by examples -

- X is an even number.
- X is a small even number.
- I usually reach to my home from my office in half an hour
- In heavy traffic it takes a long time to reach railway station.

The first example denotes a crisp possibilistic restriction, because a crisp rule (division by two) defines the set of even numbers. The second example is a fuzzy possibilistic restriction because being small integer is a fuzzy characteristic and it is a matter of graded membership in the set of small even number. The third example shows how possibilistic restriction represents information. The term usually denoted fuzziness. Similarly the last example denotes information stated in natural language through possibilistic restriction.

Finding a suitable possibilistic distribution to represent the restriction is known as precisiation of the restriction. Human mind has a remarkable ability to precisiate possibilistic restriction and make decisions on the available partial information, a process known as Natural Language Computation[16].We claim that to capture semantic information from any data, first we have to convert it into a possibilistic distribution, hence a restriction and on the basis of this distribution, decisions can be made. This method is suitable for human - machine hybrid systems, artificial intelligence and various other fields in which data, information and meaning all are important.

III. New Method to Measure Information

As it is clear from above discussion restrictions defined in natural language conveys information with meaning (semantic information), however formulation of a mechanical system which will read data and convert it into semantic information is difficult due to following reasons -

- Fitting a possibility distribution on empirical data is not easy as fitting a probability distribution.
- The precisiation of a restriction is tough when the data set is given.
- For a system generating data and information continuously (any man - machine system), defining a restriction in natural language is easy but its precisiation is not so easy.
- The process of precisiation of restriction is equivalent to finding a less specific possibility distribution that is represented by the restriction, however there are no formal techniques to select a less specific possibility distribution or to precisiate the restriction.

In view of above problems we propose a new measure of information through re-defining the possibility function on the basis of mode. A mode is the observation which has the maximum frequency. The modal element has a unique property that it is both the most possible element and most probable element of a population. In following argument we will define measure based on mode and prove that it satisfies the conditions of a possibility distribution.

Let there be a man – machine system with an observation set $X = \{x_1, x_2, \dots x_n\}$ and corresponding frequency set denoted by $F = \{f_1, f_2, \dots f_n\}$. Let the set X be uni-modular, with x_M be the mode and f_M be the modular frequency. We define the probability of an observation x_r by –

$$P(x_r) = \frac{f_r}{N} \qquad \text{where} \quad N = \sum_{i=1}^{n} f_i \tag{3}$$

We also define a map $\pi_M : (X, F) \to [0,1]$, by $\pi(x_r) = \dfrac{f_r}{f_M}$ (4)

Theorem1 - We claim that the measure defined by equation (4) is a possibility distribution for observation set $X = \{x_1, x_2, \dots x_n\}$ and frequency set denoted by $F = \{f_1, f_2, \dots f_n\}$, under the following conditions

1. Observation set X is uni-modular, where (x_M, f_M) is the modal observation and modal frequency pair.
2. The modal frequency is relatively high from any other frequency i.e. $f_M >>> f_i$.

Proof - We shall show that equation (4) satisfies the properties of possibility distribution -

Nullity – If for any observation x_j the frequency is zero $f_j = 0$, then by definition $\pi_M(x_j) = 0$.

Subset – If for observation pair x_i and x_j, $f_i \le f_j$, then by definition $\pi_M(x_i) \le \pi_M(x_j)$.

Normality – By definition $\pi_M(x_M) = {f_M}/{f_M} = 1$, thus for atleast one observation x_M, $\pi_M(x_M) = 1$.

Probability – Possibility relation – Since $f_M < N\left(= \sum_{i=1}^{n} f_i\right)$, therefore ${f_i}/{N} \le {f_i}/{f_M}$, and hence we can write $\pi_M(x_i) \le P(x_i)$ (*what is probable, that is possible).*

Maximality – Let f_{ij} denotes the frequency of $x_i \vee x_j$, while frequency of $x_i \wedge x_j$ is zero.

By definition we can write $\pi_M(x_i \vee x_j) = \dfrac{f_{ij}}{f_M}$

Now by set theory $f_{ij} = f_i + f_j$, provided frequency of $x_i \wedge x_j$ is zero, hence above equation is changed into

$$\pi_M(x_i \vee x_j) = \frac{f_i + f_j}{f_M} = \frac{f_i}{f_M} + \frac{f_j}{f_M}$$

Now by condition (2) , the modal frequency is very high than any other frequency, $f_M >>> f_i$, for any other $(x_i, f_i) \in (X, F)$. Therefore in above equation the smallest of the ratios ${f_i}/{f_M}$ or ${f_j}/{f_M}$ becomes negligible. Thus in this condition the maximum of $\pi_M(x_i)$ or $\pi_M(x_j)$ will be significant, hence we can write

$$\pi_M(x_i \vee x_j) = \max\{\pi_M(x_i), \pi_M(x_j)\}, \text{ while } \pi_M(x_i \wedge x_j) = 0 \tag{6}$$

This concludes our proof. ◊

3.1 Significance of the Condition $f_M >>> f_i$

Mode is the most possible and most probable observation of any data set. Hence for any system or scenario which generates continuous data or information, one can expect the modal element to be in any sample. Thus the modal frequency will become higher and higher and relatively the frequencies of other observations will be small, hence the condition $f_M >>> f_i$. Also in communication, particularly in human communication often information is propagated in unimodular fashion. Therefore Data-Frequency pair (X, F) becomes a possibility distribution defined by equation (4), with probability distribution defined by equation (3). The Data-Frequency pair (X, F), works as what Zadeh termed an explanatory database (ED)[19,20]. The ED is a set which contains values of the variable which helps in precisiation of the restriction or the possibility distribution. For this process, we define the following -

Given observation set $X = \{x_1, x_2 \dots x_n\}$ and its frequency set $F = \{f_1, f_2 \dots f_n\}$, be the constituents of ED (X, F).

Let the probability and possibility duo are defined by equations (3) and (4) respectively. We calculate

$$\sum_{i=1}^{n} \pi_M(x_i) = \pi_M(x_1) + \pi_M(x_2) + \dots + \pi_M(x_M) + \dots + \pi_M(x_n)$$
$$= \{\pi_M(x_1) + \pi_M(x_2) + \dots + \pi_M(x_n)\} + \pi_M(x_M)$$

Now by normality condition $\pi_M(x_M) = f_M / f_M = 1$, above equation changes into –

$$\sum_{i=1}^{n} \pi_M(x_i) = \{\pi_M(x_1) + \pi_M(x_2) + \dots + \pi_M(x_n)\} + 1$$

By the property of probability $\sum_{i=1}^{n} P(x_i) = 1$, hence -

$$\sum_{i=1}^{n} \pi_M(x_i) = \{\pi_M(x_1) + \pi_M(x_2) + \dots + \pi_M(x_n)\} + \sum_{i=1}^{n} P(x_i)$$

Or $$\sum_{i=1}^{n} \pi_M(x_i) - \sum_{i=1}^{n} P(x_i) = \{\pi_M(x_1) + \pi_M(x_2) + \dots + \pi_M(x_n)\}$$
$$= f_1/f_M + f_2/f_M + \dots + f_n/f_M$$
$$= \frac{f_1 + f_2 + \dots + f_n}{f_M}$$, where the sum does not contains f_M

Let us denote the sum $f_1 + f_2 + \dots + f_n$ by $\overline{N}$, i.e. $N - f_M = \overline{N}$ hence we get –

$$\sum_{i=1}^{n} \pi_M(x_i) - \sum_{i=1}^{n} P(x_i) = \frac{\overline{N}}{f_M} \qquad (7)$$

The L.H.S. of above equation is the difference between a possibility distribution and a probability distribution for an ED (*X*, *F*), and the R.H.S. gives us the difference between"What is possible?" and "What is probable?". Since the RHS is always positive (being the sum of positive values), this result also reconfirms that $P(Ai) \le \pi(Ai)$.Thus equation (7) becomes a measure of relative vagueness and uncertainty associated with an ED (*X*, *F*).

3.2 Proposed Method

In view of above discussion and Zadeh's theory of RCT [19], we propose the following novel method to measure information through restriction -

Step - I - To measure the information, we construct the set *X* based on the focal element(s). For example, in driving a car the focal element or observation will be traffic density or/and speed. There can be more than one focal element and hence the set *X* can be multi- dimensional.

Step - II - Next, for given observation set *X* we define Explanatory Database (*X*, *F*), consisting of continuous observation set and their respective frequencies *F*.

Step - III - We define the measures by equations (3), (4) and the difference by equation (7).These equations help us to precisiate the restriction and fit a possibility distribution on the given ED (*X*, *F*).

The composition of (X, F) changes as the system progresses through time. For example the observations which are older can be removed and then new observations are added. Thus the problem of fitting a possibility distribution on a continuous stream of data can be solved by fitting a possibility distribution ΠM and a possibility restriction precisiated by equation (7).

IV. Conclusion and Further Work

In above discussion we have proposed a method to measure information and meaning of a continuous data stream system. Various existing methods lack on the capability of approximate reasoning and decision making when information is either vague or insufficient. The method is very akin to the process of human mind analysing sensory inputs which are continuous and more often never sufficient. The method is extension of Zadeh's newly proposed Information principle (Information = Restriction), and it uses the restriction by mode (highest frequency observation). In further work we shall establish other results on possibility distribution and find representation of multi-mode systems.

References

[1]. D. Dubois and H. Prade. Fuzzy sets and statistical data. European Journal of Operational Research, 25:345–356, 1986.
[2]. D. Dubois and H. Prade. Fuzzy sets, probability and measurement. European Journal of Operational Research, 40:135–154, 1989.
[3]. D. Dubois and H. Prade. Fuzzy set and possibility theory - based methods in artificial intelligence. Artificial Intelligence, 148:1–9, 2003.
[4]. D. Dubois and H. Prade. Practical methods of constructing possibility distributions. International Journal of Intelligent Systems, 31:215–239, 2016.
[5]. D. Dubois, H. Prade, and R. Sabbadin. Decision theoretic foundation of qualitative possibility theory. European Journal of Operational Research, 128:459–478, 2001.
[6]. M. Higashi and G.J. Klir. Measures of uncertainty and information based on possibility distribution. International Journal of General Systems, 9:43–58, 1983.
[7]. G. Shafer. A Mathematical Theory of Evidence. Princeton University Press, N.J., 1976.
[8]. C. Shannon and W. Weaver. The Mathematical Theory of Communication. University of Illinois Press, Champaingn, IL, 1964.
[9]. R. R. Yager. On the completion of qualitative possibility measures. IEEE Transactions on Fuzzy Systems, 1(3):184–194, 1993.
[10]. R. R. Yager. Uncertainty representation using fuzzy measures. IEEE Transactions on Systems Man and Cybernetics, 32(1):13–20, 2002.
[11]. R.R. Yager. A foundation for a theory of possibility. Journal of Cybernetics, 10:1–3, 1980.
[12]. L. A. Zadeh. A calculus of fuzzy restrictions. In Fuzzy Sets and their Applications to Cognitive and Decision Processes. Academic Press, 1975.
[13]. L. A. Zadeh. Fuzzy sets as a basis for theory of possibility. Fuzzy Sets and Systems, 1:3–28, 1978.
[14]. L. A. Zadeh. A new direction in ai. AI Magazine, pages 73–84, 2001.
[15]. L. A. Zadeh. Towards a perception - based theory of probabilistic reasoning with imprecise probabilities. Journal of Statistical Planning and inference, 105:233–264, 2002.
[16]. L. A. Zadeh. Precisiated natural language (pnl). AI Magazine, pages 74–91, 2004.
[17]. L. A. Zadeh. Genralized theory of uncertainty: Principal concepts and ideas. Computational Statistics and Data Analysis, 51:15–46, 2006.
[18]. L. A. Zadeh. Towards extended fuzzy logic - a first step. Fuzzy Sets and Systems, 180:3175–3181, 2009.
[19]. L.A. Zadeh. Towards a restriction centered theory of truth and meaning. Information Sciences, 248:1–14, 2013.
[20]. L.A. Zadeh. The information principle. Information Sciences, 294:540–549, 2015.

Jayesh Karanjgaonkar . "On a Novel Method to Measure Semantic Information through Possibilistic Restrictions." IOSR Journal of Mathematics (IOSR-JM) , vol. 13, no. 5, 2017, pp. 32–36.