

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Generalized Information Theory: A Review and Outlook

Yong Wang

School of Computer Science and Engineering,
Guilin University of Electronic Technology, Guilin 541004, China

Abstract: Information theory was firstly used in communication. With the development of information technology, information theory has exposed its limitations and some generalized information theories were proposed. This study presents a review and discussion of the literature on the generalized information theories. Some viewpoints about the theory are proposed. An outlook of the development of generalized information theory is given. More attention should be put on the reliability and completeness of information. The multiple uncertainties of information cannot be ignored under more general conditions, especially the randomness of probability. We can naturally generalize Shannon's information theory by taking probability as random variable. That may result in a fusion of information theory, Artificial Intelligence (AI), information fusion and other relevant theories.

Key words: Probability theory, generalized information theory, uncertain, reliability, completeness

INTRODUCTION

Based on the contributions of Nyquist (1928) and Hartley (1928) on communication and information, Shannon (1948) established information theory and it is successfully used in communication and some information processing areas. As the fundamental problem of communication is that of reproducing at one point a message selected at another point and Shannon's theory faced the problem of communication, the meaning of message can be discarded. Therefore in order to form and simplify the problem, Shannon (1948) discarded the meaning of message. The theory is not universal in all the area where information technology used. Shannon (1956) pointed out that the theory was aimed in a very specific direction and not a panacea. The concept of entropy is so fundamental and multifunctional in information theory, which provides a valid measure of information under communication, so it has been abused in some theories without in-depth digestion. Some scholars realized the limitations of Shannon's information theory. Shannon's entropy is aimed at the simple communication problems and does not measure information completely, such as the divergence of information. More than 30 measures of entropies have been introduced in the literature on information theory generalizing Shannon's entropy. Classical sets must satisfy two requirements. First, members of each set must be distinguishable from one another; and second, for any given set and any given object, it must be possible to determine whether the object is, or is not, a member of the set. It was proposed the research of information theory should face a broader conception of uncertainty-based information and liberate

from the confines of classical set theory and probability theory (Higashi and Klir, 1982, 1983; Hohle, 1982; Yager, 1983). Weaver pointed out that information had three levels (Shannon and Warren, 1963): Level A, how accurately can the symbols of communication be transmitted; Level B, how precisely do the transmitted symbols convey the desired meaning; Level C, how effectively does the received meaning affect conduct in the desired way. It is obvious that Shannon's theory just settles the problems of level A. The information under this theory is called uncertainty-based information. Information theory is not explicitly concerned with the semantic and pragmatic aspects of information viewed in the broader sense (Cherry, 1957; Junarie, 1986, 1990; Kornwachs and Jacoby, 1996). But obviously these aspects are very important. The lack of consideration of these aspects certainly restricts the applicability of information theory. However, an argument can be made (Dretske, 1981, 1983) that the notion of uncertainty-based information is sufficiently rich as a basis for additional treatment, through which the broader concept of information, pertaining to human communication and cognition, can adequately be formalized. Due to the limitations of information theory, some kinds of generalized information theories were proposed. Considering the semantics and pragmatics of information, Klir (1991, 2003), Yixin (1988) and Lu (1993) proposed different generalized information theories. In the early 1990s Generalized Information Theory (GIT) was introduced to name a research program whose objective was to develop a broader treatment of uncertainty-based information. The above theories focus on such uncertainty as randomness, fuzziness, roughness and do

not consider the reliability of information. But we find even after discarding the fuzziness and roughness of information, the expression of information is also not universally appropriate. In the probability theory and information theory, the probability is always taken as fixed value, but not random variable, in reality, the probability distribution may be uncertain too. What's more, there are more uncertainty about the probability and information. In this paper, we analyze the multiple uncertainty of information from the angle of reliability. The problem of information reliability and multiple uncertainties of information was pointed out recently (Wang, 2008; Wang *et al.*, 2009a, b). This study will give a review of the Generalized Information Theory (GIT) and outlook of the new direction of generalization of information theory.

DEVELOPMENT OF MEASURE OF INFORMATION AND DIVERGENCE

As a generalization of the uncertainty theory based on the notion of possibility (Hartley, 1928), information theory consider the uncertainty of randomness perfectly. The concept of Shannon (1948) entropy is the central concept of information theory. Sometimes this measure is referred to as the measure of uncertainty. The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness or uncertainty. Shannon's model used the formalized language of the classical set theory, so it is only suitable to be used in limitation of classical set theory.

Kolmogorov (1950) proposed the notion of ϵ -entropy to measure the uncertainty when the set has unlimited elements. As pointed out by Renyi (1961) in his fundamental paper on generalized information measures, in other sort of problems other quantities may serve just as well, or even better, as measures of information. This should be supported either by their operational significance or by a set of natural postulates characterizing them, or, preferably, by both. Thus the idea of generalized entropies arises in the literature. It started with Renyi (1961) who characterized a scalar parametric entropy as entropy of order, which includes Shannon entropy as a limiting case.

As to the divergence and inaccuracy of information, Kullback and Leibler (1951) studied a measure of information from statistical aspects of view, involving two probability distributions associated with the same experiment, calling discrimination function, later different authors named as cross entropy, relative information, etc. It is a non-symmetric measure of the difference between two probability distributions P and Q . At the same time they also developed the idea of the Harold (1946) invariant, famous as J -divergence. Kerridge (1961) studied

a different kind of measure, calling inaccuracy measure, involving again two probability distributions. Sibson (1969) studied another divergence measure involving two probability distributions, using mainly the concavity property of Shannon's entropy, calling information radius. Later, Burbea and Rao (1982a, b) studied extensively the information radius and its parametric generalization, calling this measure as Jensen difference divergence measure. Taneja (1995) studied a new measure of divergence and its two parametric generalizations involving two probability distributions based on arithmetic and geometric mean inequality.

Sharma and Taneja (1977) and Santanna and Taneja (1985) studied trigonometric entropies from different aspects. The idea of weighted entropies started by Belis and Guaisu (1968). Later Picard (1979) extended it for generalized measures. After Renyi (1961), other researchers such as Havrda and Charvat (1967), Arimoto (1971), Sharma and Mittal (1975) etc., interested towards other kinds of expressions generalizing Shannon's entropy. Taneja (1989) unified some of these. Taneja (1995) introduced a new divergence measure called arithmetic-geometric mean divergence measure. Taneja (2005) studied symmetric and nonsymmetric divergence measures and their generalizations based on different divergence measures.

GENERALIZED INFORMATION THEORY

As Shannon's theory and most of the above theories are based on the classical set theory, fuzzy set and other sets, imprecise probability, semantic and pragmatic aspects of information are not considered. De Luca and Termini (1972) proposed a method to measure the uncertainty of a fuzzy set. Klir (1991) introduced Generalized Information Theory (GIT) to name a research program whose objective was to develop a broader treatment of uncertainty-based information, not restricted to the classical notions of uncertainty. In GIT, the primary concept is uncertainty and information is defined in terms of uncertainty reduction.

The various nonclassical uncertainty theories in GIT are obtained by expanding the conceptual framework upon which the classical theories are based. GIT is an outcome of two important generalizations in mathematics that emerged in the second half of the 20th century. The expansion to information theory is two-dimensional. In one dimension, the formalized language of the classical set theory is expanded to a more expressive language of fuzzy set theory, where further distinctions are based on various special types of fuzzy sets (Klir and Yuan, 1995). In the other dimension, the classical (additive) measures theory (Halmos, 1950) is expanded to a less restrictive fuzzy measure theory (Wang and Klir, 1992), within which

further distinctions are made by using fuzzy measures with various special properties. This expanded conceptual framework is a broad base for formulating and developing various theories of imprecise probabilities.

There are many results emerging from this subject. They include:

- Several nonclassical theories of uncertainty were developed. These are theories based on generalized (graded) possibility measures, Sugeno λ -measures, Choquet capacities of order 8 and reachable interval-valued probability distributions (Choquet, 1954; Gupta *et al.*, 1977)
- The above theories were generalized to a more general theory of imprecise probabilities which were based on a pair of dual measures-lower and upper probabilities and may also be represented in terms of closed and convex sets of probability distributions or functions obtained by the Möbius transform of lower or upper probabilities. Some common properties of the theories were recognized and utilized (Klir and Wierman, 1999)
- The Hartley and Shannon functions for measuring the uncertainty in classical theories of uncertainty have been generalized not only to the special theories listed above but also to other theories dealing with imprecise probabilities (Klir, 2003)
- Only some limited efforts have been made thus far to fuzzify the various uncertainty theories. They include a fuzzification of classical probabilities to fuzzy events, a fuzzification of the theory based on reachable interval-valued probability distributions, several distinct fuzzifications of the Dempster-Shafer theory and the fuzzy-set interpretation of the theory of graded possibilities (De Campos *et al.*, 1994; Tanaka *et al.*, 2004; Weichselberger, 2000; Weichselberger and Pohlman, 1990; Shafer, 1985; Klir, 2006)
- Some limited results have been obtained for formulating and using the principles of minimum and maximum uncertainty within the various nonclassical uncertainty theories. Two new principles emerged from GIT: the principle of requisite generalization and the principle of uncertainty invariance. The principle of requisite generalization requires no unnecessary limitation. the principle of uncertainty invariance requires that the amount of uncertainty (and the associated uncertainty based information) should be preserved in each transformation from one uncertainty theory to another. The primary use of this principle is to approximate in a meaningful way uncertainty formalized in a given theory by a formalization in a theory that is less general (Klir, 2006)

Lu (1993) introduced another generalized information theory. He proposed the notion of subjective information whose distinguishability is limited, including metrical information (such as conveyed by thermometers), sensory information (such as conveyed by color vision) and semantic information (such as conveyed by weather forecasts). He considers that the generalized information is subjective information that is not so precise and Shannon information is objective information whose distinguishability is unlimited. In his theory, if two weather forecasters always provide opposite forecasts and one is always correct and another is always incorrect. They convey the same objective information, but the different subjective information. The more precise and the more unexpected a forecast is, the more information it conveys. If subjective forecast always conforms with objective facts then the generalized information measure will be equivalent to Shannon's information measure, that is, the subjective mutual information equals to objective mutual information. Lu's generalized information model reflects the following ideas: when information is from forecasting, the amount of information should be inspected by fact; the more certain is the event which has been subjectively thought to be uncertain is forecasted to be (and the forecasting is right), the more the amount of information is, otherwise the amount is little or even negative. In his theory, general information is forecasting information and other information is particular case of forecasting information. Subjective information is less than or equal to objective (Shannon's) information. The generalized communication model is consistent with Popper's model of knowledge evolution. The generalized information criterion can be used to assess and optimize the pattern recognition, predictions and detection.

In Lu's theory, amount of semantics information equal to prior particularity (of the event the information referred to) minus posterior particularity. Here, particularity is a notion similar to uncertainty, it has relation to the dissimilarity and occasionality.

Lu discussed rate-of-limiting-errors and its similarity to complexity-distortion based on Kolmogorov's complexity theory and improved the rate-distortion theory into the rate-fidelity theory by replacing Shannon's distortion with his subjective mutual information. It is proved that both the rate-distortion function and the rate-fidelity function are equivalent to a rate-of-limiting-errors function with a group of fuzzy sets as limiting condition and can be expressed by a formula of generalized mutual information for lossy coding, or by a formula of generalized entropy for lossless coding.

COMPREHENSIVE INFORMATION THEORY

Comprehensive information theory was proposed to measure the syntactics, semantics and pragmatics of

information by Zhong (2002). As the name indicates, the goal of comprehensive information theory is to provide a comprehensive knowledge of every aspect about the object and its motion concerned, not only the knowledge about the form of the object's state (manner) but also the knowledge about the meaning and utility of the object's state (manner) with respect to the subject. The comprehensive information has three levels, namely syntactical information, semantical information and pragmatical information, respectively concerning the form of the state (manner), the meaning of the state (manner) and the utility of the state (manner) with respect to the subject. A unified definition of information called comprehensive information definition was given by Yixin (1988). He proposed the measure methods of syntactic, semantics and pragmatics information and then gained a comprehensive information measure formula. It was proved that some important information measure formulas are particular cases of that formula.

But Tian-Wei *et al.* (2005) pointed out that there were some problems unsettled in Zhong's theory:

- How to measure the facticity degree of state logic and effectiveness degree of state, both of them are parameters of the theory
- Is it suitable that all the information measure formulas are similar to Shannon's entropy formula?
- Is information have internal structure? What's the relative of the component elements?

Lu Chenguang pointed out there is counter-example that get unsuitable measure result from this theory in 2005.

DEFINITION OF INFORMATION

The opinions are widely divided about definition of information. Now there are over 80 kinds of definitions of information according to imperfect statistics, scholars are still discussing what the proper definition of information is and the definition of information is thought to be a puzzle. Nearly all these definitions are lacking in consideration of the reliability of information, some definitions regard information as the absolutely correct reflections of truths, no definition of information connects information with reliability (Wang, 2009).

There are many definitions of information. The following is representative ones:

- A message received and understood that reduces the recipient's uncertainty
- Knowledge acquired through study or experience or instruction

- (Communication theory) a numerical measure of the uncertainty of an outcome
- A collection of facts from which conclusions may be drawn
- The act of informing, or communicating knowledge or intelligence
- News, advice, or knowledge, communicated by others or obtained by personal study and investigation
- Intelligence; knowledge derived from reading, observation, or instruction
- Information is the objective things that exist in the world
- Information is the universal property
- Yixin (1988) gave an all-round information definition. He classified information into information on the subject level and information on cognition level, he defined information on the subject level as the self-description of the state of motion and variation pattern of the event, also defined information on the cognition level as the state of motion and variation pattern of the event that the subject percepts or expresses, including the form, meaning and utilization of the motion or pattern. He named the information on cognition level that included the form, meaning and utilization of the motion or pattern as comprehensive information (Yixin, 1988)

The above definitions does not involve whether the information correctly reflects the object the information refers to, but most information is unreliable or incomplete. In information theory, uncertainty is conventionally viewed as a manifestation of some information deficiency, while information is viewed as the capacity to reduce uncertainty. However, even when information is complete, it may be uncertain. What we want is often not the certainty but the reliability of information. The reduction of uncertainty follows the reliability. We can illustrate this problem as follows: Because of the strict discipline in one school, the time most students go to class is comparatively certain and the probability of students being late for classes is 0.01. But student A got a message from student B that student C acts worst at the inobservance of discipline (including being late for classes). It reduces the uncertainty, as far as what message that A has got from B or what about C that B has told A. But actually we can speculate that the time of attending classes of C is likely comparatively certain, according to the known situation that the discipline in the school is very strict. The prior probability of being not late for classes of B is 0.99; the probability of being late for classes of B is 0.01. But according to the message from B, the posterior probability of being late for classes of C

decreases (assume that the posterior probability of being late for classes of C is greater than 0.01). According to the calculating formula of information and the upper convexity of the entropy functions, after A got the message about C from B, if the probability of being late for classes of C is between 0.99 and 0.01, now we take the message from B for the condition, the amount of information of whether C was late does not increase but decrease and the corresponding uncertainty of whether C was late also increases. Under this condition, if people have to give a choice between the prior probability and the posterior probability after knowing the message from B, they will choose the posterior probability unhesitatingly though the uncertainty of the prior probability is lower, because the posterior probability is obtained under more complete conditions and it is more reliable (Wang *et al.*, 2009b).

In most cases, the event or information is certain, so the definition information is to reduce uncertainty is right and significative. When information is incomplete or unreliable, it is usually uncertain. Even when the information is complete and reliable, information in reality can be uncertainty as Uncertainty Principle tells us, for example, quantum state may be uncertain. The reduction of uncertainty is not suitable for these cases. If the real event is uncertain, we would rather get the uncertain result of the event than a certain one, because mostly the goal of information is to ensure facticity and reliability. If the goal is to reduce uncertainty, we can easily choose the most possible result or any of the possible results to meet this goal.

A simple definition of information was given similar to Shannon's definition: (reliable or useful) information is that to enhance the reliability. The full-scale definition shows information is the thing we get from the fact, conditions, knowledge and others on the basis of reliability under restricted conditions (such as code length limitation, calculation capacity constraints, the resolution limitation, etc.) and limited costs (Wang, 2008). When we do not know what truth is, all kinds of information that seems reliable may help us know the truth more reliable in average. Wang *et al.* (2009b) have pointed out that most information is unreliable or incomplete. Then we should consider these factors in generalized information theory.

In most cases, we want real information, but the information is usually unreliable and incomplete. If information is completely unreliable, it is useless. The definition based on reliability may make a change of direction about the research of information from single uncertainty to reliability.

RELIABILITY OF INFORMATION

As is mentioned above, the reliability of information is very important. When information is obtained, we usually investigate or judge how reliable it is. But in information theory, the reliability of information is not considered enough. What's more, the expression of information does not tell us how reliable it is. Wang (2008) found the reliability of information had relation to the multiple uncertainties of information, that is, the uncertainty of the probabilities that express the information. The more unreliable the information is, the more uncertain the probabilities are. We can give an example to explain that: Two men will fight tomorrow, but if we know nothing about the two men, we are asked to guess the probability that one will win, we may have to get the result that the probability is 0.5 as an expedience. But if we know all the states of the two men are the same by lucky coincidence, we may get the same result from the angle of information expression in information theory, that is, the probability that one will win is 0.5 as an expedience. But obviously the latter result is more useful, the prior is almost useless. That means the expression of information is deficient. If we know nothing about the two men, the probability 0.5 is very uncertain and unreliable, the real probability may be more or less than 0.5, but the more we know about the two men, the probability may be more certain and more centralized near 0.5 and hence the result is more reliable and complete.

If we consider the probability distribution of the probabilities that express the information, then Shannon's information theory may be extended to a generalized information theory that can research the reliability of information. Indeed, Shannon's information theory can be used for reference: (1) the uncertainty of probability has relation to reliability, so uncertainty theory can be used in the generalized information theory, (2) error correction coding theory can be analogously used to enhance the reliability and (3) posterior probability is more reliable and complete than prior probability, so interrelated theories can be used.

Although, reliability of information has relation to the multiple uncertainties of information, research of the reliability of information is more complex than that of uncertainty of information, as multiple uncertainties is restricted by more conditions. If we consider the reliability of information, the information theory can be fused with artificial intelligence, information fusion and other theories about information technology.

Shannon's communication system model aims at communication systems. When the model is used to realistic information system, it has limitations: (1)

Messages from information source are all thought to be entirely reliable and complete and only the reliability from the information source to the information sink are considered. (2) The information theory only considered simple channel communication, even though it considered the series and parallel of communication channels, it only considered some simple situations and in these situations, because of the stability and because each of the probability in the channel matrix is a certain value, they can be merged into a single channel. But information system in reality may be more complex, the probability in the channel matrix may be uncertain. (3) Shannon's information theory did not consider the situation that the information provides by the information source does not meet the demand of the information sink. (4) In the information theory, the information that gets by information sink is decided by the information from information source. (5) In the information theory, what the information sink wants is the information sent by the information source, but in reality the information sink may need various information.

Wang and Wang (2009) pointed out the similarities between the Shannon's communication system model and realistic information system, especially the consideration of the reliability of information. Based on the reliability of information, a generalized information system model that is suitable widely in most realistic information system and can fuse information was proposed (Wang and Wang, 2009).

In some cases, information is unilateral or incomplete, we give an example to explain the problem: When investigating a crime, we want to know who the criminal is and the suspect is only one between A and B. And we have known some information about the criminal, but this information is independent of each other. The first condition G shows that the probability that A is the criminal is c, the second condition H indicates that the probability that A is the criminal is d. And there is no overlapping existing in two conditions. When we know only one condition, we may temporarily use the probability c or d as the probability that A is the criminal by expediency. But when we know both of the conditions, we should gain more complete and reasonable probability value that the criminal is A according to the above information. In probability theory, we use conditional probability to settle the problem, but conditional probability is usually unknown, as we have pointed out, the conditional probability may be not a fixed value, but random variable or more complex variable. Under that condition, this question should be settled by a compromise between the two unilateral probabilities, an algorithm to fuse the information and enhance completeness under this condition was given (Wang *et al.*, 2009b).

OUTLOOK OF GENERALIZED INFORMATION THEORY

As we have found, Shannon's expression of information is not complete and its application is limited in communication and similar problems. That just expresses one-dimensional freedom of information, but in reality, there may be unlimited dimensional freedom. Regardless fuzziness, roughness and other uncertainties, there may be unlimited multiple random uncertainties, as we have pointed out, the probability has its randomness, parameters are needed to express the randomness of the probability, further the parameters may be random variable (or more complex variable), more parameters are needed to express the randomness, then we go to an endless loop. In most cases, the value is random variable or more complex variable rather than fixed value and fixed value is just special case of variable. Then the randomness is never completely expressed, only under some special case. As information is expressed by a set and corresponding probability distribution of the elements of the set. Shannon's model is based on a fixed set, but in reality, the set may be uncertain, even the number of elements of the set may be uncertain too. If we consider all the freedom of information, such as multiple randomness, fuzziness and roughness, uncertainty of set and etc., it is impossible to express information. Therefore, we should simplify the problem in order to establish a theory that gives consideration to the facticity (that is, to reserve the freedom of the problem without limitations or to keep the problem wholly intact) and conveniences of research (that is, be easy to express, model, theorize and operate).

Some of the above generalized theories tried to involve reliability of information, such as relative information, Lu's generalized information theory. But in these theories, there is a probability distribution that represents the true distribution of data, observations, or a precise calculated theoretical distribution and we use the true distribution to compare with the measured distribution we gained in an unreliable way, which typically represents a theory, model, description, or approximation of the true distribution. Indeed, in most cases, the truth is unknown, we should try to estimate what the truth is (to say the least, what the truth mostly is), so the problem exists, if we know the truth, the problem is settled. Lu used the probability gained from the occurred facts to inspect the gained subject probability distribution, but as we have discussed above, the probability gained from the occurred facts is just like sampling inspection, it is uncertain itself, how can we use it to inspect whether a probability distribution is reliable. As we know the fact is that the sole criterion of truth. But it can only use to prove a truth is wrong and it is not

suitable to the uncertain problems. In most cases, the truth or fact is unknown, there is no inspection standard, so under that case, what we can do is to compromise and fuse all information that is not absolutely reliable or complete. From another perspective, fact can inspect the information about itself absolutely reliably, for example, the fact today is sunny can inspect that the information today is rainy is right, but it cannot inspect the information of the weather of tomorrow or the weather of the same day of other years absolutely reliably, for we cannot get the conclusion tomorrow is sunny absolutely reliably from the reliable fact today is sunny. If every piece of information about something is not absolutely reliable, we can just fuse them to more reliable information as a band-aid.

Maybe somebody cheats and disseminates mendacious information. But when we don't know whether he is a cheat, we judge the reliability of that information to the greatest extent expediently by statistics or estimation. We suppose two cases. Under case A, though the disseminator is a cheat in reality, we are uncertain whether the disseminator is a cheat. Under another case B it is uncertain whether the disseminator is a cheat. In case A, when no more information is gained, we have no alternative but to take case A as case B. But if we know he is a cheat, we would reverse his words. When receiving information, we also try to guess or make clear how reliable it is. If information source is credible or other evidence proves it is credible, we would think it is reliable. If information is from a dishonest man, we would think it is unreliable, even inverse.

Shannon's information is so successful for its application, but the above theories seem to be not as good as Shannon's. When considering semantics of information, entropy and logarithm should be gingerly used, especially in the semantics and pragmatics aspects. Entropy has important applications in communication and file compression. In these cases, semantics and pragmatics of information is irrespective. Fuzzy sets usually concerns meaning of the information. In some theories, information is measured by one parameter even when different types of uncertainty are considered. It should make clear whether it is proper to measure uncertainty of different types (such as randomness and fuzziness) uniformly by a parameter. The single dimension expression of uncertainty is not enough. Multiple uncertainties of information may be more suitable to be expressed with several parameters in multi-dimensions. Therefore we should modestly consider whether it is proper to mix uncertainties of different types or different levels together. For example, two sample surveys on the same product get the same result, the percent of pass of the product is 0.7, but the two surveys with different

amount of samples. As information is expressed by a set and corresponding probability distribution of the elements of the set, so we do not find difference between the two surveys, but intuitively, we think the survey with large samples is more reliable. We can assimilate them with solid and flexible matters, the information more reliable or complete is more solid and infusive when different information is fused. That means the uncertainty expression of information is deficient. The uncertainties of different levels can not be easily mixed and fused to one, they express different aspects of information. We do want to know not only the probability, but also the probability distribution of that probability, that shows how reliable the information is.

Entropy is useful to measure the single uncertainty of randomness, redundancy of message and the utmost limit of compression. But when other uncertainties are considered, entropy does not provide enough measure to information. The difference, reliability and completeness of information cannot be seen from entropy. Information has the characteristics of polysemy and reciprocity, from there is rainstorm today we can get a lot of information, for example, there is disaster, there may be casualties and there may be famine next year. These results can be gained by corresponding conditional probabilities. We have pointed out conditional probability may be uncertain. The tasks of generalized information theory should include how to express these reciprocity, how to express the reliability and completeness of information, how to fuse different unreliable or incomplete information to a more reliable and complete information. Some of these tasks are now settled or to be settled by artificial intelligence and information fusion. A unification and fusion of generalized information theory, artificial intelligence and information fusion is needed.

Shannon's entropy aimed at the problem of communication, it can be used to compress data and measure uncertainty of randomness and had been proved to be right under its premise. But some measures of information just imitated form of Shannon entropy, but didn't clearly consider their application and rationality. Indeed the measure of information is a complex problem and more parameters besides entropy are needed, entropy is very useful, but not almighty.

CONCLUSIONS

In this study, we summarize the progress in generalized information theory. Typical theories are reviewed and discussed. Uncertainty is an important concept in information, but we should be aware of the multiple uncertainties of information and distinguish different levels and types of uncertainty. The importance

of information reliability is pointed out, multiple uncertainties of information involves the information reliability. Mathematical expression of a problem is often coupled with simplification of the problem and hence restricts the freedom of the problem. That causes limitations of the mathematical expression. Due to complexity, the reliability of information is hard to research. Therefore, we should make a suitable simplification to maintain the balance between facticity and conveniences of research.

ACKNOWLEDGMENT

This study was supported by Science and Research Foundation of Guangxi Ministry of Education (No. 200911MS88).

REFERENCES

- Arimoto, S., 1971. Information-theoretic considerations on estimation problems. *Inform. Control*, 19: 181-190.
- Belis, M. and S. Guiasu, 1968. Qualitative-quantitative measure of information in cybernetic system. *IEEE Trans. Inform. Theory*, 14: 593-594.
- Burbea, J. and C.R. Rao, 1982a. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *J. Multi. Anal.*, 12: 575-596.
- Burbea, J. and C.R. Rao, 1982b. On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inform. Theory*, 28: 489-495.
- Cherry, C., 1957. *On Human Communication*. MIT Press, Cambridge, MA.
- Choquet, G., 1954. Theory of capacities. *Annales de l'Institut Fourier*, 5: 131-295.
- De Campos, L.M., J.F. Huete and S. Moral, 1994. Probability intervals: A tool for uncertain reasoning. *Int. J. Uncertainty Fuzziness Knowl. Based Syst.*, 2: 167-196.
- De Luca, A. and S. Termini, 1972. A definition of nonprobabilistic entropy in the setting of fuzzy sets theory. *Inform. Control*, 20: 301-312.
- Dretske, F.I., 1981. *Knowledge and the Flow of Information*. MIT Press, Cambridge, MA.
- Dretske, F.I., 1983. *Precis of knowledge and the flow of information*. *Behav. Brain Sci.*, 6: 55-63.
- Gupta, M.M., G.N. Saridis and B.R. Gaines, 1977. Fuzzy Measures and Fuzzy Integrals: A Survey. In: *Fuzzy Automata and Decision Processes*, Gupta, M.M. (Eds.). North-Holland, Amsterdam and New York, pp: 89-102.
- Halmos, P.R., 1950. *Measure Theory*. D. Van Nostrand, Princeton, New Jersey.
- Harold, J., 1946. An invariant form for the prior probability in estimation problems. *Proc. R. Soc.*, 186: 453-461.
- Hartley, R.V.L., 1928. Transmission of information. *Bell Syst. Tech. J.*, 7: 535-563.
- Havrda, J. and F. Charvat, 1967. Quantification method of classification processes. Concept of structural α -entropy. *Kybernetika*, 3: 30-30.
- Higashi, M. and G.J. Klir, 1982. Measures of uncertainty and information based on possibility distributions. *Int. J. Gen. Syst.*, 9: 43-58.
- Higashi, M. and G.J. Klir, 1983. On the notion of distance representing information closeness: Possibility and probability distributions. *Int. J. Gen. Syst.*, 9: 103-115.
- Hohle, U., 1982. Entropy with respect to plausibility measures. *Proceedings of the 12th IEEE International Symposium on Multiple-Valued Logic, (MVL'82)*, IEEE. International, Paris, pp: 167-169.
- Jumarie, G., 1986. *Subjectivity, Information Systems: Introduction to a Theory of Relativistic Cybernetics*. Gordon and Breech, New York, ISBN: 2881240119, pp: 356.
- Jumarie, G., 1990. *Relative Information: Theories and Applications*. Springer-Verlag, New York, ISBN: 354051905X, pp: 258.
- Kerridge, D.F., 1961. Inaccuracy and Inference. *J. R. Statist. Soc. Ser. B*, 23: 184-194.
- Klir, G.J., 1991. Generalized information theory. *Fuzzy Sets Syst.*, 40: 127-142.
- Klir, G.J. and B. Yuan, 1995. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. 1st Edn., Prentice Hall PTR., New Jersey, USA., ISBN-10: 013-1011715.
- Klir, G.J. and M.J. Wierman, 1999. *Uncertainty-Based Information: Elements of Generalized Information Theory*. 2nd Edn., Physica-Verlag/Springer-Verlag, Heidelberg and New York.
- Klir, G.J., 2003. An update on generalized information theory. *Proceedings of the 3rd International Symposium on Imprecise Probabilities and Their Applications, (ISIPTA'03)*, Carleton Scientific, Lugano, Switzerland, pp: 321-334.
- Klir, G.J., 2006. *Uncertainty and Information: Foundations of Generalized Information Theory*. John Wiley and Sons Inc., New Jersey.
- Kolmogorov, A.N., 1950. *Foundations of the Theory of Probability*. Chelsea, New York.
- Kornwachs, K. and K. Jacoby, 1996. *Information: New Questions to a Multidisciplinary Concept*. Akademie Verlag, Berlin.
- Kullback, S. and R.A. Leibler, 1951. On information and sufficiency. *Ann. Math. Statist.*, 22: 76-86.

- Lu, C.G., 1993. A Generalized Information Theory. China Science and Technology University Press, Hefei.
- Nyquist, H., 1928. Certain topics in telegraph transmission theory. *AIEE Trans.*, 47: 617-644.
- Picard, C.F., 1979. Weighted probabilistic information measures. *J. Inform. Syst. Sci.*, 4: 343-356.
- Renyi, A., 1961. On Measures of entropy and information. *Proceedings of the 4th Berkeley Symposia Mathematics Statistics and Probability Statistics and Probability, (BSMSPSP'61)*, University of California Press, pp: 547-561.
- Santanna, A.P. and I.J. Taneja, 1985. Trigonometric entropies, jensen difference divergence measures and error bounds. *Inform. Sci.*, 35: 145-156.
- Shafer, G., 1985. Belief Functions and Possibility Measures. In: *Analysis of Fuzzy Information*, Bezdek, J.C. (Ed.). CRC Press, Boca Raton, FL.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27: 379-423.
- Shannon, C.E., 1956. The bandwagon. *IEEE Trans. Inform. Theor.*, 2: 3-3.
- Shannon, F.P. and W. Weaver, 1963. *The Mathematical Theory of Communication*. University Illinois Press, Urbana, pp: 117.
- Sharma, B.D. and D.P. Mittal, 1975. New nonadditive measures of inaccuracy. *J. Math. Sci.*, 10: 122-133.
- Sharma, B.D. and I.J. Taneja, 1977. Three generalized additive measures of entropy. *Elec. Inform. Kybern.*, 13: 419-433.
- Sibson, R., 1969. Information radius. *Probability Theor. Related Fields*, 14: 149-160.
- Tanaka, H., K. Sugihara and Y. Maeda, 2004. Non-additive measures by interval probability functions. *Inform. Sci.*, 164: 209-227.
- Taneja, I.J., 1989. On Generalized Information Measures and their Applications. In: *Advances in Imaging and Electron Physics*, Hawkes, P.W. (Ed.). Academic Press, New York, pp: 327-413.
- Taneja, I.J., 1995. New Developments in Generalized Information Measures. In: *Advances in Imaging and Electron Physics*, Hawkes, P.W. (Ed.). Academic Press, New York, pp: 37-135.
- Taneja, I.J., 2005. On Symmetric and Nonsymmetric Divergence Measures and Their Generalizations. In: *Advances in Imaging and Electron Physics*, Hawkes, P.W. (Ed.). Academic Press, New York, pp: 177-250.
- Tian-Wei, W., W. Huanchen and Z. Xu, 2005. A survey of information measurement methods. *Syst. Eng. Theor. Methodol. Appl.*, 14: 481-486.
- Wang, Y., 2008. Analysis and betterment of shannon's information definition. *J. Inform.*, 27: 57-60.
- Wang, Y., 2009. Analyses on limitations of information theory. *Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence, (AICT'09)*, IEEE International, Shanghai, pp: 85-88.
- Wang, Y. and H. Wang, 2009. A generalized information system model. *Proceedings of the 3rd International Conference on Genetic and Evolutionary Computing*, Oct. 14-17, IEEE Computer Society, Guilin, China, pp: 154-157.
- Wang, Y., H. Wang and J. Huang, 2009a. Information fusion algorithm enhancing information completeness. *Proceedings of the 3rd International Conference on Genetic and Evolutionary Computing*, Oct. 14-17, Computer Society, Guilin, China, pp: 407-409.
- Wang, Y., H. Wang and X. Tang, 2009b. On the reliability of information. *Proceedings of the Chinese Control and Decision Conference*, June 17-19, IEEE Press, pp: 871-874.
- Wang, Z. and G.J. Klir, 1992. *Fuzzy Measure Theory*. Plenum Press, New York.
- Weichselberger, K. and S. Pohlmann, 1990. *A Methodology for Uncertainty in Knowledge-Based Systems*. Springer-Verlag, New York, pp: 132.
- Weichselberger, K., 2000. The theory of interval-probability as a unifying concept for uncertainty. *Int. J. Approximate Reasoning*, 24: 149-170.
- Yager, R.R., 1983. Entropy and specificity in a mathematical theory of evidence. *Int. J. Gen. Syst.*, 9: 249-260.
- Yixin, Z., 1988. *Principles of Information Science*. Fujian Press, China.
- Zhong, Y., 2002. *Information Science Principle*. 3rd Edn., Beijing University of Posts and Telecommunications Press, Beijing.