The study on the out-structure information classification based on observers

Fan Li¹, Zhisen Wang*¹
¹ Controlling Science & Engineering, Dalian Polytechnic University, Dalian, Liaoning, 116034, China
*Corresponding author’s e-mail: lifandlpu@foxmail.com

Abstract. Information classification is imperative for machine identification in artificial intelligence. Current information classification approaches only satisfy the mechanization of machines for human services. An essential question is still unresolved, which motivates people to find a way to make machine work for human with thinking. This thesis first discusses information classification in a broad perspective, using a method different from traditional ones which only pay attention to technology and ignore integrity. Then it proposes the three dimension information model of "physical entities - information entities - observers". In this paper, we divide information into “natural information” and “social information”. And the process of out-structuring information is divided into “generalized communication” and “generalized transmission”, which puts forward new ideas for the generalization and systematization of information research.

1. Introduction
The rapid development of society is largely attributed to the fast circulation and effective use of social resources. Generally speaking, social resources can be divided into material and information. Accurate and detailed classification is the precondition of resource recycling and exploitation. The emergence of the Internet of things has shown many shortcomings of the existing machine classification algorithms, posing new challenges to the information research, and pushing people to re-examine information from a new perspective.

Threshold is the standard of classification. If we make life information as threshold, physical entities can be divided into living entities and inanimate entities. Because of the being of the sub-threshold, the two categories of entities can also be divided into different kinds. There has been a great deal of researches on the relationship between information and physical. This paper is based on the assumption that "material and information is mirror symmetry" is right, and the physical entities’ classification can influence the classification of information. Because information has many special characteristics such as virtual etc., its classification method is different from physical entities.

Generalized information is no longer confined to the "0", "1" bits stored on the Internet. Everything can be information, according to the difference among time, space and observes, objects forms various classes and kinds. The way information transmitted between substances varies by categories. In this paper, we propose the two concepts “generalized communication (GC)” and “generalized transmission (GT)”. Communication of same kinds between information source and sink is defined as GC. Otherwise it is called GT, whether the purpose of communication is reached or not. It is basic and important for studying GC and GT that information can be classified based on different observers, which is also the scalar to measure the degree of industrialization.
Research on information has been going a long time, however, there has been very little research on the classification of information. And the limited research on information classification is based on big data [5], the Internet [6], computer technology and related classification algorithms [7]. Although these views have made great contributions to the development of information science and information society, it does not solve the problems of information isolated islands, information fragments, and information explosion etc., which are now in conflict with the great integration of society. New research ideas and methods are urgently needed to break the bottleneck of information development.

2. Basic problems research

2.1 The status quo of information research

Big data, the Internet, and Internet of things are in full swing, information flattening breaks the original information pyramid structure. Classical information classification mostly belongs to two kinds of objective information based on ontology and the subjective information based on epistemology. Objective information emphasizes that the existence of information has nothing to do with human existence. The subjective information emphasizes the influence and plays the important role on information [8]. Ontological objective information solves the problem of information existence, but it does not explain the emergence, transmission and use of information. Subjective information pays attention to the importance of people. The authors think that the two views are limit, because both of them do not jumped out of the category of philosophical information research, which is not correspond to generalized information.

The universality of artificial intelligence [9] has made machine learning [10] become one of the most popular vocabulary at present. Deep learning [11] is a branch of machine learning, and it is has been used in many domains, such as image recognition, nature language processing, machine translation, etc. Accurate and comprehensive information classification will be benefit for improving the speed and accuracy of machine learning, and paves the way for the realization of man-machine governance in the future digital society.

Physical entities tend to be ordered, organized and logical based on their detailed classification. On the contrary, information entities are fragmented and disorganized in nature. Physical entities are mapped to be a very different image in information space, which form another space: digital space [12]. Except human beings, the primary subjects of digital society also include information entities. Efficient information classification makes it possible to "cross boundary" [13]. In summary, it is necessary to study observer-based information classification.

2.2 Methodology

**Proposition 1:** Everything can be an event with time $t$.

Comprehensive study is the basic for generalized information research. This paper makes comprehensive analysis for information based on different kinds of physical entities and various observers. In this methods, the authors use logical reasoning [14] in mathematics to establish the correlation among physical entity - information entity - observers - out-structure information (o.i) – outer - structure information space, which makes the research of information theory more general and universal.

Material is objective and real, it can be described by $f(x, y, z)$, $(x, y, z)$ denotes spatial position, $t$ is time. Static physical entities are conceptual hierarchies, we use $f(x, y, z)$ describe one of the static physical entities. However if time $t$ is added, it will become events, denoted by $f_d(x, y, z, t)$, this process is shown in equation (1).

\[
\begin{align*}
&f : f_s(x, y, z) \rightarrow f_d(x, y, z, t) \\
&\text{(1)}
\end{align*}
\]
Proposition 2: The relation between physical entities and the information entities is mirror symmetric.

According to proposition 2, we can get that both physical and information entities have objective existence attributes. Mapping between the two can be expressed as equation (2).

\[ g : g_p \xrightarrow{a(observe)} g_i \] (2)

Where, \( g_p \) and \( g_i \) denote physical entities and information entities respectively. According to equation (1) and equation (2), we can also get that \( g_p \) and \( f_d(x, y, z, t) \) satisfy equation (3).

\[ g_d \subseteq f_d(x, y, z, t) \] (3)

The same reason, mapping between \( f_s \) and \( g_i \) can be described by equation (4).

\[ f^o g : f_s(x, y, z) \xrightarrow{t o observe} f_i(x, y, z, t) \xrightarrow{a(observe)} g_i \] (4)

As shown in equation (4), we can obtain that mapping among physical-time-space can be realized by adding time and observer.

3. Formation of the out-structure information

3.1 Physical entities classification

In this paper, according to the characteristics of the physical entities, and referring to the chemical and biological classification of substances, physical entities are divided into living physical entities and inanimate physical entities, as shown in Table 1.

| Physical Entities          | Inanimate Physical Entities | Living Physical Entities |
|---------------------------|-----------------------------|--------------------------|
| Inanimate entities        | Artificial inanimate entities | Animals | Plants | Microorganism | Human beings |
| that exist naturally      |                             |             |         |              |               |

3.1.1 Features of physical entities. Inanimate physical entities: their structures are relatively simple, and the morphology of these entities is low, such as mountains, rivers, flowers, trees, and so on.

Living physical entities: The physical structure of these entities is relatively complex. Its form is higher, and the order of the individual is high. According to Shannon and Wiener's research on information [15, 16], we can get that these entities contain more information, whose level is higher.

3.2 Observer classification

Substance and information are objects of observation and study. According to above introduction and proposition 3, observers can be divided into human beings, animals, plants, and artificiality. Table 2 presents characteristics of different observers.

3.3 Formation Process of Out-structure Information

In this section, we will introduce GC and GP. The formation of out-structure information should experience the process of coding, delivery, reception, and storage. Different observers and observation entities correspond to different methods.

If observers and observation entities are in the same category, the form of information out-structure is GC. Otherwise, the process is called GP. Information is encoded in the same form among the same kinds, and communication from source to sink through the media can be realized in accordance with the agreed manner. The encoding of information flows is different among different kinds. When they enter the media, it cannot follow the unified encoding and decoding methods to transfer information between source and sink, so media only broadcasts different information streams between source and sink. Communication and propagation are two different forms of information exchanging.
Table 2. Comparison of four different observers

| Classification | Property       | O.I means      | O.I mode                  | Sociality | O.I capability |
|----------------|----------------|----------------|---------------------------|-----------|----------------|
| plant          | Living entities| Sensory organs | Pollen, gene, etc.       | individuals | low           |
|                |                |                | Voice, body language, heredity, etc. | groups     | middle         |
| animal         | Living entities| Sensory organs |                           |            |                |
|                |                |                |                           |            |                |
| artificiality  | Inanimate entities | Intelligent devices | Language, words, electrical signal, etc. | nothing     | low           |
| Human being    | Living entities| Sensory organs |                           |            |                |
|                |                | Intelligent devices | Voice, language, words, picture, heredity |            |                |

Next we will discuss the relation between narrow sense communication and generalized communication.

The narrow communication process can be described as equation (5).

\[ r(t) = s(t)\xi(t) + n(t) \] (5)

Where, \( r(t) \) denotes the received information by the base station. \( s(t) \) is information transmitted by the source. \( \xi(t) \) denotes attenuation factor of channel. \( n(t) \) is the noise in the transmission process. Equation (5) can be extended to express the process of generalized out-structure information, as shown in equation (6).

\[ R(t) = S(t)\xi'(t) + N(t) \] (6)

Where, \( R(t) \) denotes the sum of generalized information. \( S(t) \) denotes the actual information contained in physical entity. \( \xi'(t) \) denotes information enhancement factor with observer. \( N(t) \) denotes information received by other observers.

The out-structure information depends on the presence of the observer; Out-structure information varies according to the observers; Out-structure information is distorted by the influence of observers.

4. Out-structure information classification based on observers

4.1 Three dimensional out-structure information model

The symmetry model between physical entities and information entities is shown in figure 1. We put observers into figure 1(a), and then get the important views of three dimensional out-structure information model, as shown in figure 1(b). Because observers and observation objects are limited, and the ability of different observers to construct information is different, which makes the out-structure information and the observed physical world asymmetrical. If we use \( I_{OE} \) denotes the total of the out-structuring information entities, we can get \( I_{OE} << I_E \), where \( I_E \) denotes total information entity.

As shown in figure 1(b), \( f(I_O) \) and \( g(p) \) are feedback coefficients, the former denotes the feedback coefficients from observation space to out-structure information space, the latter denotes the feedback coefficients from out-structure information space to space observation space. This model can be considered as a circle system that is made of a series of feedback. The quality of the system output depends on whether the feedback is accurate or not. Observer is an important factor in the feedback coefficient, the capability of generate information of observers reflects the accuracy of the feedback.
Suppose that we examine observers and observation entities from more than one dimensional, we’ll find that the received information will be closer to the observed entity itself, and less distortion.

4.2 Out-structure information classification
In this section we will discuss information classification from two levels of nature and society. The universe is a highly integrated, dynamic, cyclic ecosystem with self-regulation, self-appreciation, self-repair, and adaptive capabilities. Because of their species level, all things in the universe can be divided into two levels. In this paper, we divide the information structure into natural structure and social structure. And the out-structure information can also be divided into two corresponding categories: natural information and social information. Figure 2 shows the hierarchical structure of information.

Natural Information: This kind of information is generally considered as structure, function, attribute, etc.. All physical entities contain natural information, and its biggest characteristic is stabilization.

Social information: This kind of information is produced by the members of the society after the emergence of social organizations. Generally speaking, social information is related to human beings, and the information can also be produced by animals and plants. Natural information is also included.

The hierarchical structure of information can change from lower structure to higher structure, but this change depends on the extent of the observer's evolution. Many existing machine identification algorithms can only recognize the simple structure of images, which hinders the development of machine identification because of the lack of analysis of social structure and emotional structure.

5. Conclusions
The spread of AI has made machine learning become one of the most popular technologies at the moment. Accurate and comprehensive information classification helps to improve the speed of machine learning, and pave the way for the realization of human-computer interaction in the future digital society. Observers are bridges between the physical world and the virtual world of information. Therefore, it is necessary to study the information classification based on different observers. According to the discussion of some basic research, we propose the following points: (1) The process of out-structuring information can be divided into two forms, which are generalized communication and generalized transmission. (2) By introducing observers into matter and information, we build physical entities - information entities - observers three dimensions information model. (3) We divide
information into natural information and social information, and they correspond to two layers of nature and social respectively.

References

[1] A. G. Billings and R. H. Moos, "Coping, stress, and social resources among adults with unipolar depression," *Journal of Personality & Social Psychology*, vol. 46, pp. 877-91, 1984.
[2] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, pp. 2787-2805, 2010.
[3] Y. Desmedt and Y. Frankel, "Threshold cryptosystems," in *Advances in Cryptology*, 1989, pp. 307-315.
[4] G. J. Klir, "Uncertainty and information: Foundations of generalized information theory," *Neural Networks IEEE Transactions on*, vol. 58, pp. 756–758, 2007.
[5] D. Klein, P. Tran-Gia, and M. Hartmann, "Big Data," *Informatik-Spektrum*, vol. 36, pp. 319-323, 2013.
[6] D. Benavraham, "Resilience of the Internet to Random Breakdowns," *Physical Review Letters*, vol. 85, pp. 4626-4628, 2000.
[7] A. Wilcox and G. Hripcsak, "Classification algorithms applied to narrative reports," *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, vol. 71, p. 455, 1999.
[8] Zhang Long, Liu Li, Du Weiwei, Liu Li, Wang Zhisen, Investigations of information communication and information space from the perspective of observe[J]. *Journal of Dalian Polytechnic University*, 2017, 36 (5): 308-385.
[9 ] R. Sangal, H. Mehta, and R. K. Bagga, "Proceedings of the 20th international joint conference on Artificial intelligence," *Japanese Journal of Applied Physics Supplement*, vol. 39, pp. 149-154, 2014.
[10] Mitchell, *Machine Learning*: McGraw-Hill, 2003.
[11] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436, 2015.
[12] D. Bertol, Designing digital space: an architect's guide to virtual reality: John Wiley and Sons Ltd., 1997.
[13] M. Tortoriello and D. Krackhardt, "Activating cross-boundary knowledge: The role of Simmelian ties in the generation of innovations," *Academy of Management Journal*, vol. 53, pp. 167-181, 2010.
[14] E. Amir and S. McIlraith, "Partition-Based Logical Reasoning," 2002, pp. 389--400.
[15] C. E. Shannon, "A mathematical theory of communication," *Bell Labs Technical Journal*, vol. 27, pp. 379-423, 2014.
[16] N. Wiener, *Cybernetics: Or Control and Communication in Animal and the Machine*: MIT Press, 2000.