

2012 AASRI Conference on Modeling, Identification and Control

The Study of Secondary Intelligent Organization Methods of CBT Training Library based on Information Entropy

Hemin Du^{*}

School of Art & Media, Xi'an Technology University, Xi'an, China

Abstract

In order to realize the intelligent of computer based training system, this paper puts forward a knowledge form retrieval method based on information entropy. Through analyze the inner link which existing between the knowledge forms, we can build knowledge base, and can find the knowledge forms which cause the failure of training through information entropy calculation. Use information entropy and combination entropy calculation formula, we can retrieval and construction the knowledge chains which cause training failure, and regroup into learners training knowledge base. Overall, this paper puts forward the algorithm of knowledge base secondary organization, this algorithm can automatic ruled out the training knowledge which the leaner has mastered, and we can improve the efficiency of CBT system through the strengthening training..

© 2012 The Authors. Published by Elsevier B.V. Open access under [CC BY-NC-ND license](#).
Selection and/or peer review under responsibility of American Applied Science Research Institute

Keywords: knowledge base, knowledge element, intelligent organization, information entropy

1. Introduction

CBT (Computer Based Training) is the product of computer technique highly development and popularization. At present, the conventional CBT system provides the rich knowledge base for student to

^{*} Corresponding author. Tel.: 0-86-13201660920
E-mail address: xtdn-dhm@126.com.

practice and test repeatedly, but it is the new hot technology for CBT research that is how to setting up reasonable training knowledge base for different training phase, especially use computer intelligent technology to eliminate the knowledge points which the student have mastered and realize the intelligent organization for training knowledge base.

The intelligent organization of knowledge base is based with the people's cognitive research. The professor Chen Qing-quan who is academician of Chinese academy of engineering summarize the current age with BINC, BINC namely biological technology, information technology, nano technology and cognitive technology^[1]. Cognitive science is interdisciplinary combination with philosophy, psychology, linguistics, anthropology, and neuroscience and computer science^[2]. Yu Sheng-quan has pointed out that the third time digital learning has already "from transmission and behavioral paradigm to the constructiveness cognitive paradigm, and then turned to the situation cognitive paradigm^[3]".

From the perspective of cognitive science, the strengthen training is the best way and important method of learning to improve memory ability and ultimately effect. For the intelligent of CBT system, the purpose is strengthen training, through evaluating the student's training effect which have finished on the prior stage of learning, the system can automatically analyze the student forgotten knowledge (namely knowledge element which lead to the training failure), search the related knowledge chain, and constructs the secondary training knowledge base, so can make the student have effective strengthen memory training with specific aim and achieve economic and efficient training purpose.

2. The intrinsic cognitive mechanism of knowledge base

The training knowledge base of CBT system is constructed by much basic knowledge form. From the view of cognitive, there are some certain order and hierarchical relationships for knowledge forms between each other in the knowledge learning and memory, and the knowledge forms link to mesh structure. The knowledge base chains relationship can be divided into order chain and support chain, order chain is the most basic relations between knowledge forms which express the progressive of information accepted, support chain express the interaction support relationship between each other, as shown in figure 1 below.

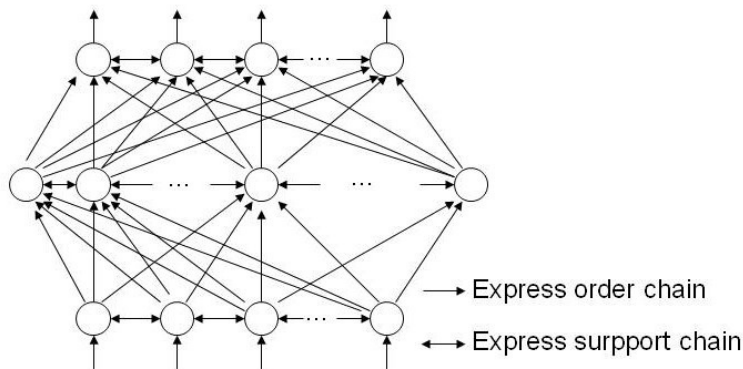


Fig. 1. the connectionism network model of information point (knowledge element)

Based on cognitive thinking and decision-making principle, the first step of training plans intelligent organization is knowledge element detection, it is a related illation realization process that the knowledge element is examined. For the knowledge element search of CBT system, it is sure that the knowledge element

which leads to the failure of training existed in the training knowledge base, so the reasoning process of knowledge element detection is a kind of typical certainty reasoning.

In the knowledge base system of CBT, we can regard it as a figure for knowledge network, it is a process which search for the knowledge in the figure that the related knowledge element is searched, the process can be form a figure search strategy, and we can express the basic figure search strategy as the flow description in figure 2 below:

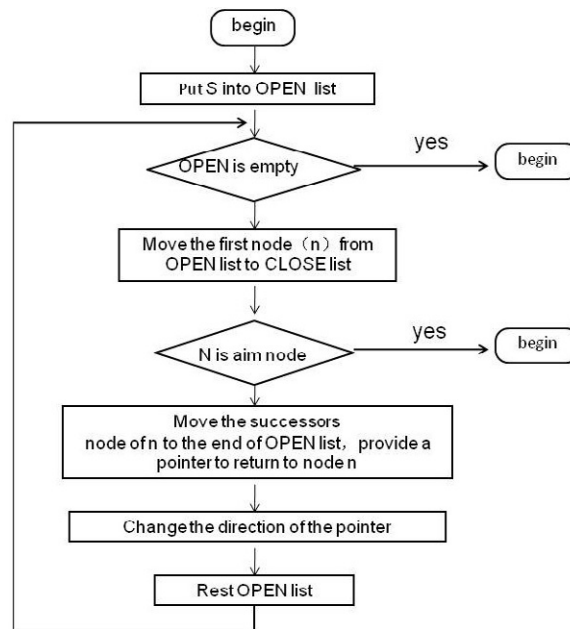


Fig. 2. reasoning process diagram of CBT knowledge element search

The secondary scheme of CBT system intelligence organization for training results which shown in figure 2 can be described as follow: the knowledge base which formed by the basic knowledge can be regarded as the whole search figure, the knowledge element storage container which retrieved and cause the failure of training can be regard as OPEN list, the eventually retrieved knowledge element storage container can be regard as CLOSE list. For the current training results, we can score points as hundred when processing intelligent illation retrieval. If the score of current training result is 100 points, it shows that the training entirely correct, and knowledge element has completely mastered, there does not exist searchable contents, namely the OPEN list is empty, search failure, end the search; If the score of current training result is not 100 points, it shows that there exist some knowledge element in OPEN list, choose the knowledge element which caused the training failure, mark it as n , and move it from OPEN list to CLOSE list. We can know from connectionism network model that there maybe exist support chain in knowledge element (show that n is a expanded node), the expanded node will also be corresponding detection, and moved to CLOSE list; Now, the rest of contents (the successor node of n , which be put at the end of the OPEN list) would be rearranged in the OPEN list and form a new OPEN list. The system would continue to illation with the above reasoning rule until the end of illation, so the search illation for knowledge element which would be all may lead to failure of training is over.

3. Knowledge base information entropy

The figure search strategy which based on cognitive reveals the reasoning principle and process which can find the cause training failure knowledge element in knowledge base. We can use it to realize the knowledge reasoning through the information entropy algorithm.

For example, in the aircraft maintenance CBT training system, it constitute a complete knowledge base by each part of the basic knowledge, assembly knowledge and fault detection knowledge, when the student detect whose learning effects, the reasons that led to the training failure are various, there are not only due to the learn reasons, but also because of some unexpected reasons. However, the system be only able to detect the knowledge element factor, in other words, it only find which knowledge element cause the training failure, because there are the complex relationship between the knowledge elements, therefore, there maybe exist the uncertainty of information target.

In the 1960s, the United States cybernetics scholars L.A.Zadeh put forward and development the fuzzy mathematics method; in the 1980s, the polish mathematician Z.P awlak put forward the rough set theory; in addition, there is gray system, and so on^[4]. Though these methods showed very good performance in uncertainty, indeterminacy and imprecision for processing information^[5], and have made great achievements in intelligent information processing, but in the perspective of CBT system knowledge reasoning, it's difficult to connect them with the retrieval of knowledge element, it is mainly because of the complexity of the joined knowledge network which make the relationship of relevant knowledge element are unlikely to quantitative description. However, Claude Elwood Shannon put forward the information entropy which shows the huge potential in this problem.

Shannon puts forward the concept of information entropy in a mathematical theory of communication in 1948, which is regarded as the foundation of modern information theory. Entropy theory establishes the mathematical model for the communication process ^[6]. Shannon's entropy theory is limited to the mathematical theory of communication system, so it is called narrow sense information theory. However, the application of information theory far more than communication theory, it has cross with physics (statistical mechanics), mathematics (probability theory), electronic engineering (communication theory) and computer science (algorithm complexity) ^[7].

4. The information entropy algorithm for secondary organization of knowledge base

In the aspect of mathematics (include probability theory and statistics), the basic concepts of the information theory which include entropy, relative entropy and mutual information can be defined as a probability distribution function, any one of them can depict and expression the behavior characteristic of long sequence random variables, them can be used to estimate the probability of rare events, and find the best of index error in the hypothesis. Based on this, we can use the concept of information entropy as the foundation of implementing knowledge element intelligent search in the CBT system construction.

4.1. Information Entropy

A system more orderly, the information entropy is lower; conversely, a system more chaos, the information entropy is higher. Therefore, from the perspective of the information theory, the information entropy is a measure degree of system orderly.

The basic concept of information entropy includes self-information, information entropy and combination entropy.

Self-information: used for information measure, when the information source sent out a signal $x_0 \in \mathcal{X}$, the provided amount of information is called of x_0 self-information, record as $I(x_0)$, it is the measurement values of uncertainty of signal x_0 . In the knowledge element searching, the probability of knowledge element x_0 which affect some training failure is recorded as $P(x_0)$. For $x \in \mathcal{X}$, the greater the probability $x \in \mathcal{X}$, the greater the chance of the occurrence, and the smaller the value uncertainty $I(x)$.

For $I(x)$, there is formula 1:

$$I(x) = c \log \frac{1}{p(x)} \quad \dots \dots \dots (1)$$

C is constant, the base number of logarithm is 2, and the unit of self-information is bit.

Information entropy: it is for the entire source, express the amount of average information for each one of these signals. If use random variable represents an information source, the entropy is its measurement of the average uncertainty. In the CBT system, set X is random variables between the value and knowledge element, the probability distribution function is $p(x) = Pr\{X = x\}$, $x \in \mathcal{X}$ having:

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x) \quad \dots \dots \dots (2)$$

Combination entropy: in the CBT system, the knowledge element which influences the training result may be transverse, multidimensional, so we import the concept of combination entropy. Set the joint distribution of n dimension random variables is $p(x_1, x_2, \dots, x_n) = Pr\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$, there is $x_1 \in \mathcal{X}^1, x_2 \in \mathcal{X}^2, \dots, x_n \in \mathcal{X}^n$, then, the formula of combination entropy is:

$$H(X_1, X_2, \dots, X_n) = - \sum_{x_1 \in \mathcal{X}^1} \sum_{x_2 \in \mathcal{X}^2} \dots \sum_{x_n \in \mathcal{X}^n} p(x_1, x_2, \dots, x_n) \log p(x_1, x_2, \dots, x_n) \quad \dots \dots \dots (3)$$

4.2. The intelligent organization algorithm of knowledge base information entropy

We can put forward the concept of knowledge base information entropy that connects knowledge base with information entropy to constitute the secondary training scheme organization for CBT system.

The basic theory of knowledge base information entropy: the all knowledge in the CBT training library constitute knowledge base in the form of knowledge element, if there is any errors or failures in training, then show that there are knowledge element is active state which related the training content, and whose information entropy tends to maximize, in other words, the more information entropy of any knowledge element, the larger influence to the training result.

The mathematics representation of knowledge base information entropy: set have k knowledge element in anyone training knowledge base. According to contain a combination mathematics knowledge element whose item amount is C_k^i (means different i items have C_k^i , $i=1, \dots, k$). So there have the possible different knowledge element (item sets) of $2^k - 1$ kinds. Set every different item sets expressed by random variable X_i ($i=1, 2, \dots, 2^k - 1$), suppose P_i express appear probability of random variables X_i (it is the i item sets), there is $\sum_{i=1}^{2^k-1} P_i = 1$, now need to evaluate the most reasonable estimate probability, we can use the maximum entropy principle, calculate the best probability which is made $H = -P_i \log P_i$ ($P_i \geq 0$ and $P_i \log P_i = 0$ (if $P_i = 0$)) get the most probability.

According to Lagrange method number, structural function:

$$F = H - \alpha \left\{ \sum_{i=1}^{2^k-1} P_i - 1 \right\} \quad \dots \dots \dots (4)$$

Calculate the partial derivative for F , get:

$$\partial F / \partial P_i = -1 - \log P_i - \alpha \quad (i=1, 2, \dots, 2^k - 1) \quad \dots \dots \dots (5)$$

Make:

$$\partial F / \partial P_i = -1 - \log P_i - \alpha = 0 \quad (i=1, 2, \dots, 2^k - 1) \dots\dots\dots (6)$$

Get:

$$\log P_i = -1 - \alpha \quad (i=1, 2, \dots, 2^k - 1) \dots\dots\dots (7)$$

It is:

$$P_1 = P_2 = \dots = P_{2^k - 1} = e^{-1-\alpha} \dots\dots\dots (8)$$

Because $\sum_{i=1}^{2^k-1} P_i = 1$ so

$$P_1 = P_2 = \dots = P_{2^k - 1} = 1 / (2^k - 1) \dots\dots\dots (9)$$

Through the above analysis, we can obtain the related definition of knowledge base information entropy:

Define: set any knowledge base A , there has k kind of different knowledge element, the k kind of knowledge element form $2^k - 1$ kind of different item sets. Set the real probability of each item is $P_i (i=1, 2, \dots, 2^k - 1)$, so:

$$I(A) = (-\sum_{i=1}^{2^k-1} P_i \log P_i) / (\sum_{i=1}^{2^k-1} \log(2^k - 1) / (2^k - 1)) = (\sum_{i=1}^{2^k-1} P_i \log P_i) / (\log 2^k - 1) \dots\dots\dots (10)$$

$I(A)$ is the information entropy of knowledge base A .

The information entropy of knowledge base $I(A)$ has the nature properties $0 \leq I(A) \leq 1$. The information entropy is smaller, express the effects of learning for knowledge base is good, and there are less that the students don't mastered knowledge element, in other words, $I(A)$ more close to 0 shows that the training effect more better, $I(A)$ more close to 1 shows that the training effect more worse, there are more active knowledge element in knowledge base, and there are larger workload which need to strengthen training, based on this, the system can automatically rearrange the knowledge element, excluding the mastered training knowledge, and the secondary training knowledge base generated automatically.

5. Conclusion

The key of secondary intelligent organization for CBT training is to look for the knowledge base which leads to the failure of training. This paper start from the perspective of cognitive science, puts forward the concept of knowledge base information entropy, and built the knowledge element search algorithm which can realize the intelligence of CBT system; it has important meaning for improving availability and efficiency of the CBT system. As the specific application test, we have developed the intelligent computer based aircraft maintenance training system which proved the feasibility of this algorithm. Now, in the aspect of the system perfection, there are some details problem (such as the division of the knowledge element particle size and weight coefficient between knowledge elements) need to be solve in the next work.

References

- [1] Information on http://www.eet-china.com/ART_8800445979_865371_NT_738ff067.HTM
- [2] M.Eysenck M. T. keane: *Cognitive Psychology* (Psychology Press, UK 2000).
- [3] S.Q.Yu. From knowledge transfer to the cognitive construction, then to situated cognition. China audiovisual education [J], (2007) No. 6, P.7 (in Chinese).
- [4] D.B.Zhang, Y.N.Wang and F.Duan. Application of rough sets to extract region of interest in intelligent empty bottle inspection [J]. Journal of system simulation, Vol.19 (2007) No.5, P.1021 (in Chinese).
- [5] Y.X.Kou, L.Wang and W.J.Liu. Application of information entropy theory to analyze combat mission

state requirement [J]. Journal of System Simulation Vol.21 (2009) No.5, P.1427 (in Chinese).

[6] C.E.Shannon. A Mathematical theory of communication [J]. The bell system technical journal, Vol. (1948) No.3, P.373.

[7] M.C.Thomas: *Joy A Thomas. Elements of Information Theory* (John Wiley & Sons, USA 2006).