Method of Cluster Based Cross-Domain Knowledge Acquisition for Biologically Inspired Design

Shen Jian, Hu Jie, Ma Jin, Peng Ying Hong, Fang Yi, Liu Wen Hai

Abstract—Biologically inspired design inspires inventions and new technologies in the field of engineering by mimicking functions, principles, and structures in the biological domain. To deal with the obstacles of cross-domain knowledge acquisition in the existing biologically inspired design process, functional semantic clustering based on functional feature semantic correlation and environmental constraint clustering composition based on environmental characteristic constraining adaptability are proposed. A knowledge cell clustering algorithm and the corresponding prototype system is developed. Finally, the effectiveness of the method is verified by the visual prosthetic device design.

Keywords—Knowledge based engineering, biologically inspired design, knowledge cell, knowledge clustering, knowledge acquisition.

I. Introduction

BIOLOGICALLY inspired design is a systematic design method of knowledge-driven design, which aims to establish a systematic cross-domain knowledge analogy method through the research and imitation of natural phenomena or biological systems. With biologically inspired design, the inspiration for problem solving and technological innovation in the engineering field is provided. With the different directions of knowledge flowing in the design process, the methods of biologically inspired design include: solution-driven and problem-driven. Solution-driven design uses the principle knowledge known in the biology domain and constructs the functional model as the bridge, which can be used as a connection of the engineering domain knowledge modeling, retrieving, fusing and analogizing, so as to obtain satisfactory engineering solution. Problem-driven innovation design takes the given engineering design issue as an outset, through analyzing functional model and understanding the design matters in the field of biological, a set of products functional requirements of the original solution is obtained [1].

The acquisition of excited source in biological field is a prerequisite for biologically inspired design, which has a decisive influence on the quality of the subsequent design activities. In the field of biology, the two main objects of this paper are to find the analogical examples corresponding to the engineering problems and to provide design inspiration for problem solving in the engineering field.

Shen Jian is with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, PR China (phone: +86-021-34206552, e-mail: shenjan0819@sjtu.edu.cn).

Hu Jie is with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, PR China (corresponding author, e-mail: hujie@sjtu.edu.cn).

At present, the methods of obtaining the examples of biological domain are divided into two types: the method of expert consultation and the method based on structured knowledge database. Case-based expert advice receiving is a means of obtaining relevant biological examples, and many conceptions are presented when consulting experts who have similar experiences. Zhao [2] analyzed the mechanism of changing octopus skin color by the contraction and expansion of the muscles that allows different quantities of light through it. The control of the electric field of polymerization tension within the object in the field of engineering is to achieve this function of simulation. Then a new direction for flexible display research is opened up. Festo [3], a Germany company, developed a bionic penguin, bionic seagulls, bionic elephant nose and jumping Kangaroo robots after investigating Bionic Design for the simulation of biological morphology and functional incentives. Although the method of expert consultation has been widely used in the field of biology, it does rely too much on the personal experience of the expert which may be derived from occasional research and imitation of natural phenomena. There is plenty of randomness and blindness when the number of experts is small, that is the reason why it is difficult to promote the use of systematic. Zhang [4] suggested a method taking the SBF model as a cornerstone to build an ontology designed to establish an instance ontology associated model in the biological domain. Based on the ontology model, a DANE prototype system was developed. Tinsley [5] used function-based biological knowledge modeling to construct a biological function knowledge database, and realized the process of converting the functional principle into the engineering system. Biomimicry Institute has collected academic literatures related to biomimetic design to create a network-based system known as, AskNature [6]. The framework of biological knowledge database based on function index is carried out. The site already has tens of thousands of examples in the biological domain. In order to settle the problem of low retrieval efficiency, Chakrabarti [7] proposed an algorithm combining internal and external clustering in combination with the SAPphIRE model. Haberland and Kim [8] presented a constraint-based framework for clustering retrieval. Vattam [9] proposed a solution-driven model based on the keyword clustering algorithm. The unstructured instance knowledge model returned by the above knowledge acquisition method is not descriptive; furthermore, it is difficult to be applied and it lacks research on cross-domain knowledge retrieval. To tackle those drawbacks, a measure based on the proposed functional knowledge meta-modeling is presented in this work, which

transforms the cross-domain example knowledge retrieval problem of the biologically inspired design process into clustering retrieval of discrete knowledge in the biology domain or engineering domain. In the line with the cross-domain terminology knowledge representation, some cross-domain search functional words are defined. A one-stage clustering of functional semantics based on functional feature semantic relevance is performed, while the classification based on different types of environmental characteristics with the combination of the biology domain functional and environmental characteristic constraints, named two-stage clustering, is completed.

II. SELECTION OF FUNCTIONAL WORD

In cross-domain knowledge representation, different terminology is often used to denote that a same function or same term has different meanings in discriminate areas [10]. For example, "bear" in the biological domain and "manufacture" in the engineering domain have the same meaning, and can be used in the function of generating new things. The term "seal" in the engineering field means the "prevent leakage" of the element, and in the biological field it refers to the "mammal animal". In the biological domain, knowledge clustering is to cluster knowledge cells with similar functions and environmental characteristics. The highest membership degree of the cluster can be achieved in the knowledge reuse process in accordance with design requirements, thus the generation of an engineering domain

solution is motivated. As a result of the characteristic of the natural language's knowledge representation in the biology domain and knowledge clustering in the cross-domain, it is difficult to determine the retrieval keywords in the early stage of the clustering search. The process of function word selection is depicted as follows:

Firstly, describing and retrieving. The most relevant function of the verb is selected to represent the design requirements of the extracted function in the knowledge base library. After the retrieval of the knowledge cell library, the relevant knowledge cell is acquired. If no search results are returned, the synonyms of the function verbs are selected from the standard verb tables to extend the search range and repeat the search until search results are returned. The related knowledge cells are retrieved and annotated, and a collection of high frequency nouns coexisted with the verb is summarized and sorted by the frequency of occurrence.

Secondly, correcting and repeating. It is determined whether the acquired knowledge cell satisfies the design requirement; otherwise, the knowledge cell library in the light of the repeated nouns is researched to obtain the related function verbs.

Thirdly, determining and obtaining. The analogous knowledge cell sets are analyzed to determine the input-output flow and environment constraints related to design requirements. The corrected functional verbs and input and output streams are used to participate in the knowledge clustering process in the biological domain.
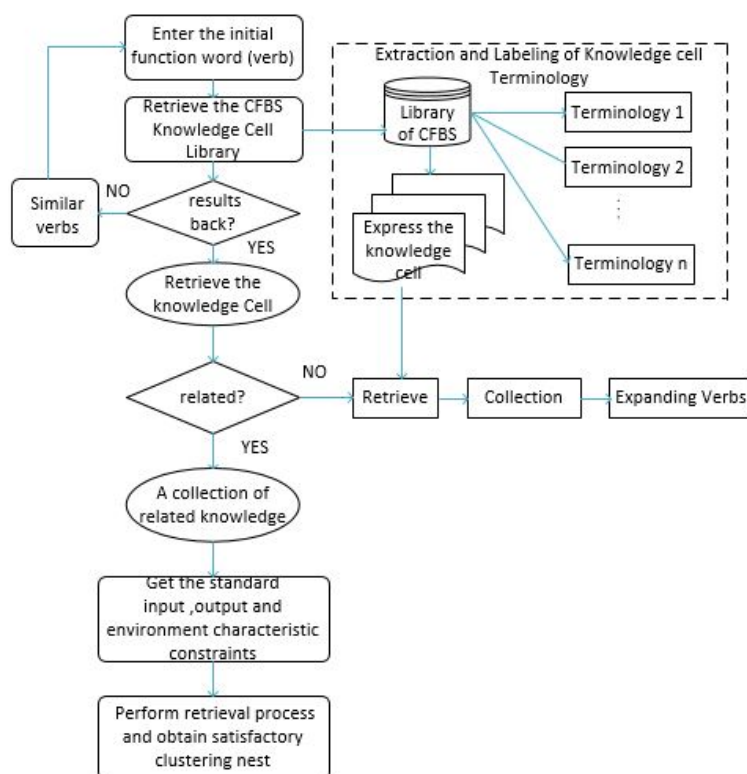

Fig. 1 Retrieval process of the related biological CFBS knowledge cell cluster

III. KNOWLEDGE METHOD IN BIOLOGICAL DOMAIN

A. One-Stage Clustering

Semantic relevance of functional features is the spatial measures of semantic similarity between functional words. The features of the semantic model with nodes state the terms and links state the relationship between terms, as shown in Fig. 2.

Step 1: A single knowledge cell function is composed of seven basic elements from the function word, and is represented as: $F_T^i = \{V, INF(S,M,E), OUF(S,M,E)\}$, where F_T^i $(i = 1,2,...,n)$ represents the knowledge function, V is function verbs in knowledge cell function names, $INF(S,M,E)$ and $OUF(S,M,E)$ represent the input and output streams of the knowledge cell, respectively. Where $T_m^i \in F_T^i, T_n^j \in F_T^j$ $(m = n = 1,2,...,7)$ represent the corresponding functional words in the functional feature representation of the two knowledge cells, respectively; its corresponding node location is shown in Fig. 1 and the calculation model is expressed as:

$$a_k = Sim_{SS}^{IC}\left(T_m^i, T_n^j\right)_k = max_{N_L \in Sup\left(T_m^i, T_n^j\right)}\left\{-\log\left(P(N_L)\right)\right\} \quad (1)$$
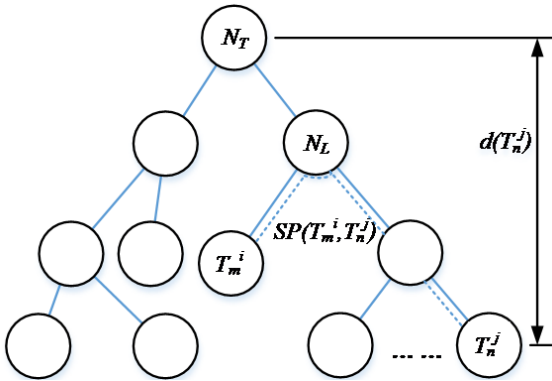

Fig. 2 Semantic model in functional semantic clustering

Where $Sim_{SS}^{IC}\left(T_m^i, T_n^j\right)_k$ is the function word $\left(T_m^i, T_n^j\right)$ of the kth $(k = m = n)$, it is the common parent node with the highest level of the two terms, and $P(N_L)$ is the probability of occurrence of all terms of N_L, the value range is $[0\ 1]$.

Step 2: The nodes of association strength with level dependent is defined as $T_{sub(n)}$, the node of T_n, and $d\left(T_{sub(n)}\right)$ is the depth from the N_T to the $T_{sub(n)}$, and the association between the adjacent nodes intensity is calculated as:

$$\omega_n\left(T_n, T_{sub(n)}\right) = 1 + 1/2^{\left(d\left(T_{sub(n)}\right)-1\right)} \quad (2)$$

The semantic relevance between functional words is affected by the relationship of affiliation. The weights of adjacent nodes are defined as follows: $same - to\ (\mu_s = 1) \succ synonymous - of\ \mu_m) \succ is - a\ (\mu_i)$, empirically: μ_m=0.86, μ_i=0.64.

While there is affiliation between T_m^i and T_n^j, $SP(T_m^i, T_n^j)_k$ is the shortest distance of T_S^i to T_T^j, the semantic correlation between nodes can be defined as:

$$ND\left(T_m^i, T_n^j\right)_k = \frac{1}{n}\sum_{n \in SP(T_m^i, T_n^j)}(1 - \mu_n) \times \omega_n\left(T_n, T_{sub(n)}\right) \quad (3)$$

In order to transform the node distance into a similarity, the inverted function is introduced. Thus, ND similarity $\left(Sim_{SS}^{ND}\left(T_m^i, T_n^j\right)\right)$ of terms T_m^i and T_n^j of the kth compared functions pair is displayed as follows:

$$b_k = Sim_{SS}^{ND}\left(T_m^i, T_n^j\right) = \left\{ND\left(T_m^i, T_n^j\right)_k + 1_k\right\}^{\wedge -1} \quad (4)$$

Step 3: In adaption with the function in the representation of knowledge cell, the same weight (ω_F=1-ω_V=0.5) is given to the compound verbs ($⟦(\omega⟧\_V$=0.5) and the input-output streams, the composite semantic relevancy expression should be:

$$c_{ij} = Sim_{\frac{IC}{ND}}\left(F_T^i, F_T^j\right) = \frac{ra_1+(1-r)b_1}{2} + \frac{\sum_{k=2}^{7}[ra_k+(1-r)b_k]}{12} \quad (5)$$

where r is the weight of information content based IC functional words. Considering the characteristic of ND and IC, the value of r is about: $[0.85\ 0.95]$.

B. Two-Stage Clustering

Generally speaking, different organisms, no matter how far apart on the phylogenetic tree, have the same function or approximate phase structure to adapt to the environment where the biological environment is the same or converging, which is known to us as convergent evaluation. In the marine environment, the function of headway (Figs. 3 (a)-(d)) is achieved when mammals such as whales and seals have generated a structure similar to fish "fins". For the purpose of obtaining outside information, arthropods (e.g., insects) have evolved an organ having the same function as a mammal (e.g., a human).

Faced with different biological environments, transformation towards different functions in different directions will occur in the same species - this is called divergent evolution. Fig. 3 shows how the desert lizard evolved heavy scales to prevent water loss (Fig. 3 (e)), the spotted lizard in desert enhanced the wing membrane to glide (Fig. 3 (f)), the serpentine lizard in the wetland environment has a water drifting function to traverse the water (Fig. 3 (g)), and the chameleon can change color to avoid predators in the rainforest (Fig. 3 (h)). The biological structure of the function and its surrounding environment are closely related.

The constraint of environment feature is composed of a series of characteristic parameters, and its algorithm of similarity framework is depicted in Fig. 4.
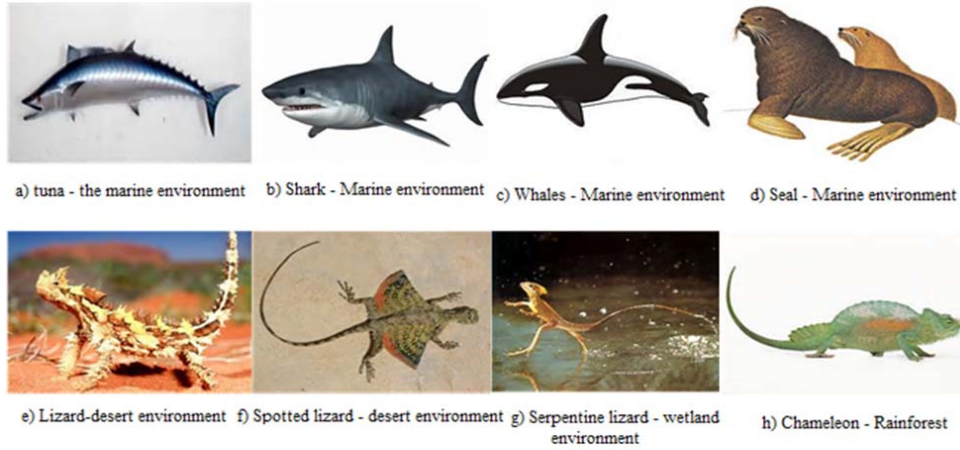
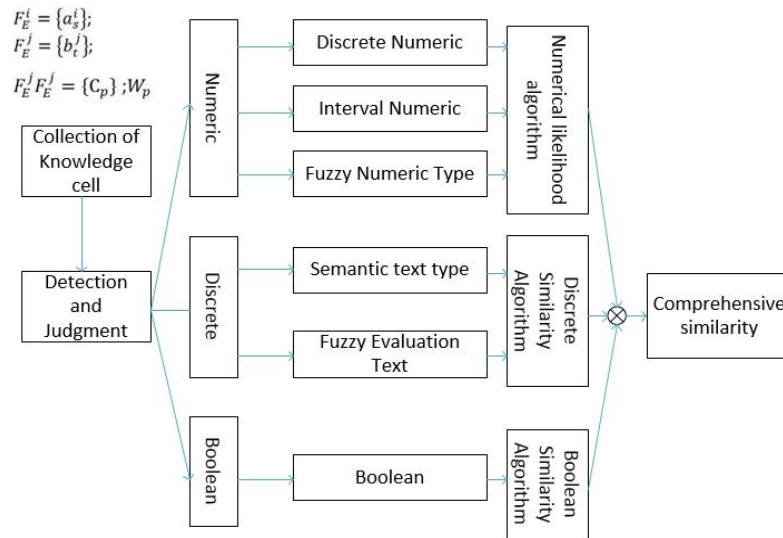Fig. 3 Environmental effect on the biological function evolution

Fig. 4 The schema of similarity calculation for environment feature constrains

With the introduction of the FCM clustering algorithm into the process of knowledge cell clustering, the AFCM algorithm is proposed and the environment constrained clustering based on the adaptability of environmental characteristics constraints is realized by combining the different types of constraints of similar environmental feature algorithms. In the preliminary clustering process, the objective function AFCM corresponded to is proposed as follows:

$$Z(P_h, C_1, C_2, \ldots, C_v) = \sum_{i=1}^{v} \sum_{j=1}^{k} u_{ij}^2 D_{ij}^2 \qquad (6)$$

where C_i is the ith fuzzy cluster, v is the number of clusters, u_{ij} is the degree of affiliation of the jth $(j = 1, 2, \ldots, k)$ knowledge cell to the ith cluster, D_{ij} is the distance between the jth knowledge cell and the ith cluster C_i.

Coordinated with the definition of membership in fuzzy set, membership function will satisfy the following condition:

$$P_h = \left\{ u_{ij} \middle| \begin{array}{l} u_{ij} \in [0\ 1]; \sum_{i=1}^{v} u_{ij} = 1; \\ \forall t; \sum_{t=1}^{k} u_{ij} < n \end{array} \right\} \qquad (7)$$

When the minimum value is achieved, the objective function $Z(P_h, C_1, C_2, \ldots, C_v)$ is defined as: $Min\{ Z(P_h, C_1, C_2, \ldots, C_v)\ s.t.\ U \}$.

In the case of optimizing objective function, the Lagrange multiplier is introduced, which finally is deduced as:

$$\sum_{i=1}^{v} u_{ij}[s] = 1 \qquad (8)$$

$$u_{ij} = [\sum_{m=1}^{v} (\frac{D_{ij}}{D_{mj}})^2]^{\wedge -1} \qquad (9)$$

In the calculation of D_{ij}, the environmental characteristics of comprehensive similarity calculation results and the corresponding degree of membership are used to calculate the distance between cluster computing; the formula is defined as:

$$D_{ij} = \left\| \frac{\Sigma_{m=1}^{v} u_{im}^2 d_{mj}^2}{\Sigma_{m=1}^{v} u_{im}^2} \right\| \qquad (10)$$

In the second-order clustering process, the initial membership function $U[0]$ is given, then the distance between the knowledge cell and cluster D_{mn} is calculated in adaption with (10). The new membership function $U[i] = 1,\ldots,s,$ is calculated continuously in accordance with the order of loop iteration, until the matrix norm of the membership matrix between the two iterations satisfies the iterative terminated criterion: $\|P_h^{(l+1)} - P_h^{(l)}\| \leq \varepsilon$. On the terminating of the iterative process, the clustering result and the membership matrix based on the environmental characteristic constrains and outputs are acquired.

IV. CLUSTERING INSTANCE

Attention has been focused on research increasingly to restore the visual perception of the blind. Retinal prosthesis, visual cortex prosthesis and optic nerve prosthesis [11]-[13] have been classified into two types on the basis of their different pathways and principles of transition of visual information. By analyzing the mechanism of the eye, it can be determined that the human visual system consists of adjusting to basic visual clarity and controlling view angle. The visual prosthetic device design requirements functional model is expressed as: visual signal acquisition → visual signal transformation → visual signal transmission → visual signal coding → coding signal transmission → encoding information control. The design requirements of the engineering field are expressed in Fig. 5.
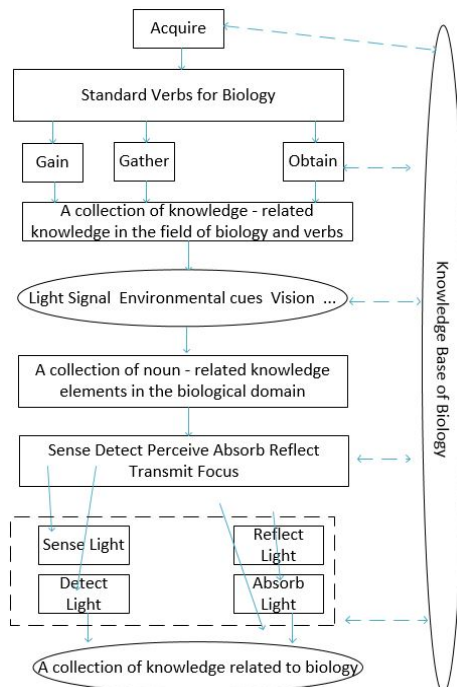

Fig. 5 The flow chart of finding the proper keywords to describe design requirements

The reference range of the visual prosthetic device design is expanded to search the structures and strategies which are corresponding to the other biological visual signal acquisition functions in nature; in this manner, the necessary design incentive for the visual prosthetic device and reference is generated. The process of cross-domain function keyword selection is shown in Fig. 5.

Considering the analytic hierarchy process, the weight of the environmental characteristic constraints is calculated by using the $1 \sim 9$ scale method proposed by Santy. The weight of the environmental characteristic constraint: "motion, eye type, illumination intensity, sight distance, angular field of view" is: $\{w_m, w_e, w_i, w_s, w_a\}$ = {0.342, 0.381, 0.039, 0.175, 0.063}.
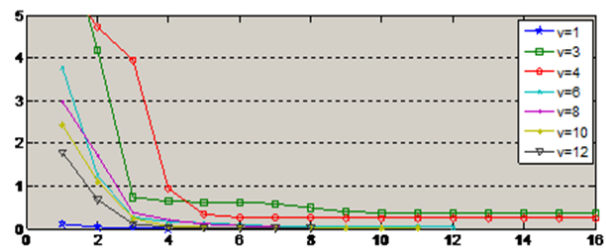
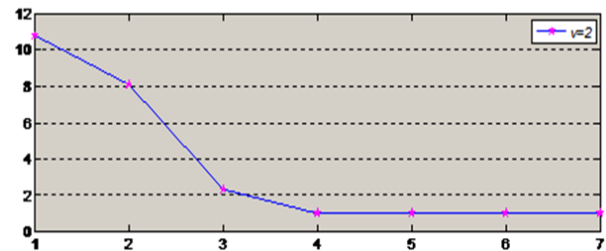Fig. 6 Environmental constraints based clustering (v=2)

Fig. 7 Objective function value chart with different cluster number

The clustering of 16 samples generated by first-order clustering is carried out. The terminated threshold of iteration is: $\varepsilon = 0.0001$. The maximum number of iterations: $r = 25$, respectively, and the floating trend of the objective function and the corresponding membership degree under different cluster number are investigated. The floating trend of the objective function $Z(P_h, C_s, A)Z(P\_h, C\_s, A)$ during clustering iteration is shown in Fig. 7.

As the number of cluster v equals to 2, the membership function corresponds with the clustering result is the cluster C1 or C2, respectively. The row vector represents the knowledge cell set from 7 to 22, and the value related with intersection point is the membership degree of the corresponding knowledge cell belonging to its corresponding cluster.

The number of clusters is set as: $v = 1 \sim 16$, the corresponding trend of the objective function is shown in Fig. 8. The results of fuzzy clustering tree are shown in Fig. 9.

The clustering results are listed as follows: C1 = {7, 8, 13, 17, 18}, C2 = 1, 2, 3, {9, 12, 14, 16, 19, 20}, C3 = {10, 11}, C4 = {15} and C5 = {21, 22}, the clustering results are indicated in Fig. 9.
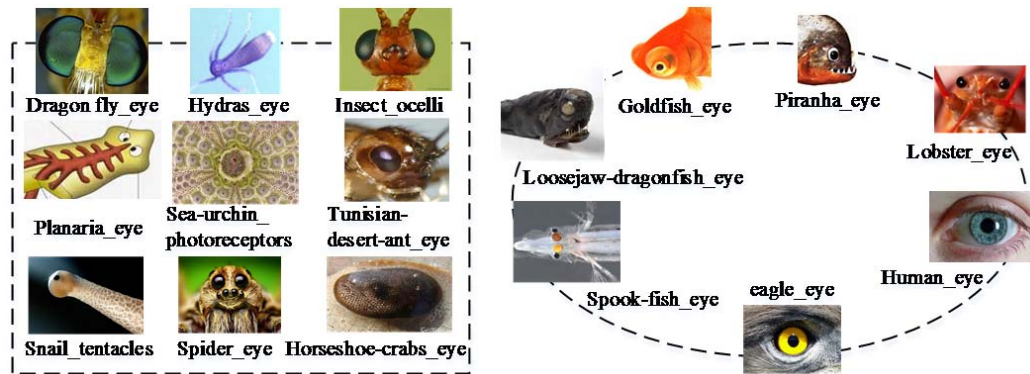

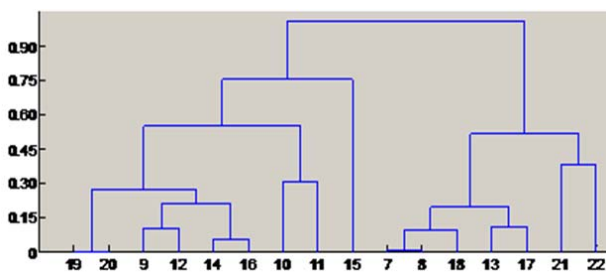Fig. 8 Objective function value chart with different cluster number

Fig. 9 Cluster tree of AFCM clustering

Taking the clustering results expressed in Fig. 10 into consideration, a conclusion can be draw that clustering sets C1 and C5 perform visual signal acquisition by Camera eye, and the muscle is used to pull the eyeball to expand the signal acquisition range. Vision is formed when signals are transmitted to the visual center and visual pathway is constructed simultaneously. The collection of C1 is aquatic, its refractive index of water is similar to intraocular transparent medium. The light is refracted into the eye and gathered to the rear of the photoreceptor site - it can be moved back and forth while C5 can change the curvature of the lens to achieve imaging and adjust the imaging distance. The clustering distance calculation in hierarchical clustering is introduced into the FCM algorithm and AFCM is proposed.

The comparison of AFCM, Hierarchical clustering algorithm and K-Means is shown in Fig. 11. It can be concluded that the average clustering accuracy of the AFCM algorithm is higher than K-Means and hierarchical clustering algorithm. The biggest difference between the distinct clustering methods is 26%. As for clustering time, however, while the sample data is small, its computation complexity is high. It makes little effect on small sample clustering process, such as biology domain or engineering domain knowledge cell clustering. Therefore, the AFCM clustering algorithm proposed in this paper can realize the demand of knowledge cell clustering retrieval in view of knowledge representation in the biology domain and engineering domain.
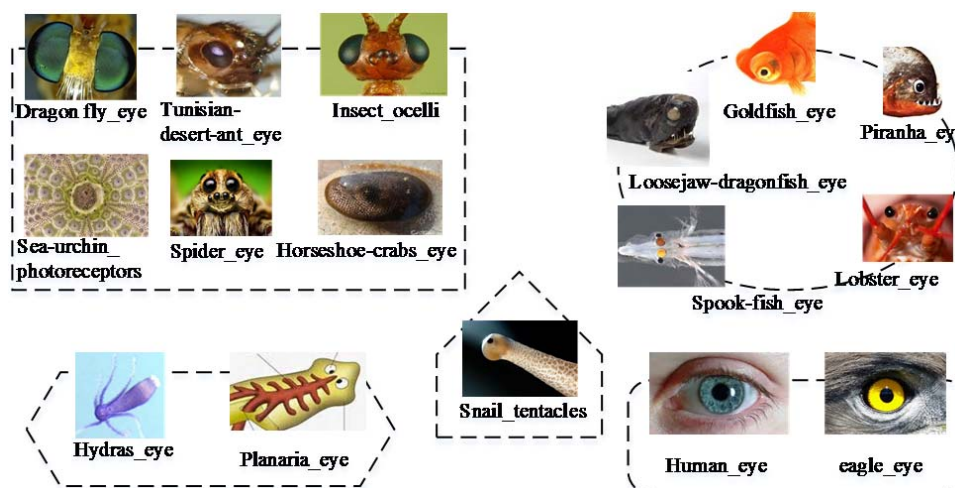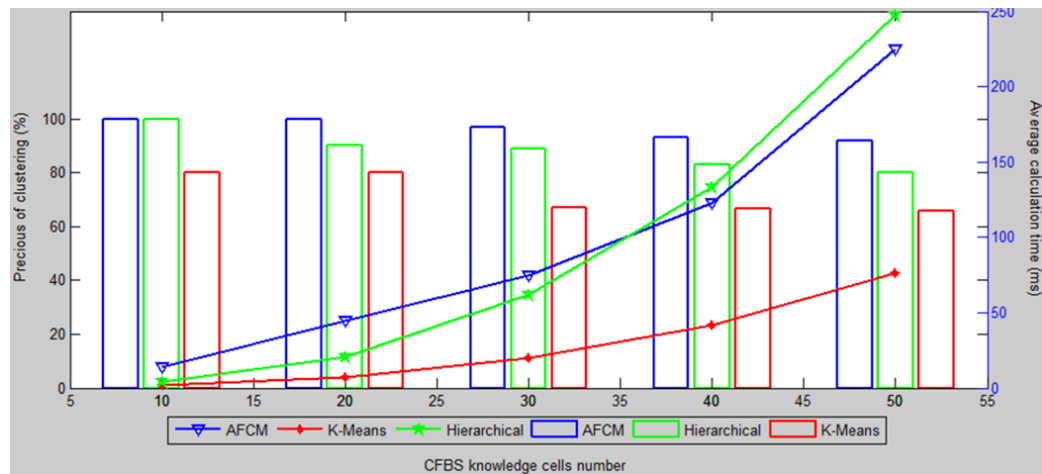

Fig. 10 Output clustering results (v=5)

Fig. 11 Comparison of AFCM, Hierarchical clustering algorithm and K-Means

V. CONCLUSIONS

There are numerous defects including contingency in acquisition, inefficiency of knowledge acquisition, which leads to the shortage of cross-domain knowledge acquisition in existing methods of biologically inspired design. With the existing knowledge meta-model, the domain instances are discretized into functional modules. Implementing one-stage clustering of functional semantics based on functional semantic relevance and two-stage clustering based on different types of environmental feature constraints and function, knowledge cells related of knowledge clustering are gathered as the biological excitation design process knowledge database. Finally, the visual prosthesis device is taken as an example. It is obviously that the clustering time, accuracy and efficiency are improved significantly on the basis of clustering results. On the one hand, the algorithm avoids the discretization of knowledge distribution across domains and reduces the number of objects in the design process effectively; while on the other hand, it can obtain the existing design knowledge reasonably full of certain practical value.

REFERENCES

[1] Nagel J. K. Systematic design of biologically-inspired engineering solutions (J). Dissertations & Theses - Gradworks, 2010.
[2] Wang Q, Gossweiler G R, Craig S L, et al. Cephalopod-inspired design of electro-mechano-chemically responsive elastomers for on-demand fluorescent patterning(J). Nature Communications, 2014, 5:4899.
[3] Bionics Ants. A project of Festo AG & Co. KG Corporate Design Ostfildern Germany (DB/OL). 2015, available from: www.festo.com/bionic.
[4] Zhang, D.Q., Chen, S., Pan, Z.S, et al. Kernel based fuzzy clustering incorporating spatial constraints for image segmentation(C). Machine Learning and Cybernetics. 2003 International Conference on. IEEE. 2003, 4: 2189-2192.
[5] Tinsley A, Midha P, Nagel R, et al. Exploring the use of functional models as a foundation for biomimetic conceptual design (J). American Society of Mechanical Engineers, 2007:79-92.
[6] AskNature, A project of the Biomimicry Institute (DB/OL). 2008-2016, available from: http://asknature.org.
[7] Chakrabarti, A. Supporting Analogical Transfer in Biologically Inspired Design (J). Biologically Inspired Design, 2014:201-220.
[8] Haberland M, Kim S. On extracting design principles from biology: I. Method-General answers to high-level design questions for bioinspired robots (J). Bioinspiration & Biomimetics, 2015, 10(1):016010.
[9] Goel A K, Vattam S, Wiltgen B, et al. Information-processing theories of biologically inspired design (M) //Biologically Inspired Design. Springer London, 2014: 127-152.
[10] Ma J, Hu J, Feng J F, et al. Constrained FBS knowledge cell model, representation, and applications for conceptual design (J). Archive Proceedings of the Institution of Mechanical Engineers Part C Journal of Mechanical Engineering Science 1989-1996 (vols 203-210), 2015, 230(11).
[11] Couzin-Frankel J. Second chapter (J). Science, 2016, 353(6303):983.
[12] Normann R A, Maynard E M, Rousche P J, et al. A neural interface for a cortical vision prosthesis (J). Vision Research, 1999, 39(15):2577-2587.
[13] Li M, Yan Y, Wu K, et al. Penetrative Optic Nerve-Based Visual Prosthesis Research(M)//Artificial Vision. Springer International Publishing, 2017: 165-176.