Acta Technica 62 (2017), No. 6B, 1061–1072

Construction of domain knowledge ontology map based on concept learning of neural networks

Ting Yu[1], Maoyi Tian[2]

Abstract. In order to improve performance enhancement for spectrum construction algorithm of domain knowledge ontology, a construction method for spectrum of domain knowledge ontology based on concept study of self-organizing neural network shall be proposed. Firstly, according to structure model for domain ontology graph, frame of ontology learning shall be set up. Then relation mapping for knowledge term-term shall be established; secondly, model for self-organizing neural network shall be adopted for construction of fuzzy clustering method to realize sorting algorithm for KLSeeker knowledge ontology of dichotomy relation. Semi-supervised learning of ontology graph shall be realized based on concept clustering to decrease human intervention; finally, effectiveness of algorithm can be verified through experiment analysis.

Key words. Self-organizing neural network, Concept learning, Knowledge ontology, Clustering analysis.

1. Introduction

Domain knowledge refers to set of concept, correlation, constraint set, state and its change rule in some domain, and it covers inferential mode and evolutionary relationship within domain, which is knowledge and content of some application and industry background. At present, researches related to domain knowledge home and abroad are mainly focused on three aspects of knowledge discovery, knowledge sharing and knowledge application, and scholars are widely concerned about domain knowledge distance (domain knowledge similarity) which is taken as core problem of knowledge discovery.

[1]School of Computer Science ,Sichuan Technology and Business University,Chengdu city of Sichuan Province, China,611745

[2]School of Applied Technology, XiHua University, Chengdu city of Sichuan Province, China, 610039

Based on hypergraph structure, Cai Shuqin et al. expressed for multi-relation knowledge, and calculation model of knowledge similarities was established; based on domain ontology thinking, Chen Jie et al. set up calculation model for initial similarity of concept domain knowledge and similarity of concept non-hyponymy relation; Yang adopted web ontology language (Web Ontology Language OWL) for description of discipline ontology, and domain knowledge distance based on attribute was established, which effectively solved searching of domain knowledge of learning resources; Yang Li et al. gave model for domain knowledge distance based on information quantity, and nearest sorting method SDkNN was established based on it; Xu Dezhi et al. and Li Wenjie et al. discussed calculation method for domain knowledge distance between concepts and model for overlapping degree of domain knowledge; Patwardhan et al. completed measurement and analysis for domain knowledge relation between concepts at the usage of word net and context vector; Douglas et al. discussed about searching and expanding technology for domain knowledge relation of word bank, and applied domain knowledge analysis to browsing and searching for word bank structure. Domain knowledge network of word bank of certain automatic characteristics and interactive control was established; distance for network domain knowledge of word bank was provided with overall review by Budanitsky et al.; based on usage of OWLS for description of service, Ge et al. also gave distance algorithm for domain knowledge between two nodes oriented at grid service; oriented at Web service and based on formalization and inference of describing logic, Gao et al. gave distance method for domain knowledge matched in contents based on introducing calculation ontology concept, and system realization under environment of knowledge visualization discovery was completed; Ofer et al. raised an easy calculation method for domain knowledge distance oriented at computerized inference , and realized distance set by taking advantage of aggregate function; Del et al. also raised measurement method for domain knowledge distance based on knowledge and oriented at organizing and searching related domain knowledge information; Kuper et al. established domain knowledge of Euclidean distance of data for geographic information system (Geographic; Information System GIS) by taking advantage of self-defined database language, and it was compared with work of Kuijpers et al. and Cisl, which shows that such method conforms to natural characteristics more.

2. Formalization expression of knowledge structure

5 tuples of knowledge ontology can be divided into 2 aspects: abstract element and formalization element. Abstract element is equivalent to connotation of knowledge, and it is made up of concept and its related information. Formalization element corresponds to extension of knowledge, and it is formalization expression of things. It is with the aid of feature of specifically describing things oriented at concept attribute assignment, and it is an instance of concept. Relation between concept and instance is shown in Fig.1.

Abstract element of knowledge ontology is core of knowledge ontology, and it is more originated from expression of human to prior knowledge. But for knowledge in a certain domain, it has already constituted a relatively completed knowledge

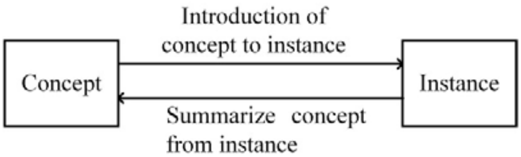


Fig. 1. Relation between concept and instance

structure. Therefore domain knowledge structure can be used to describe abstract element of knowledge ontology at the time of discussing knowledge ontology. In knowledge structure, knowledge element can be used to express concept, and proving mechanism of knowledge correlation and knowledge inference shall be established. Their relation is shown in Fig.2.

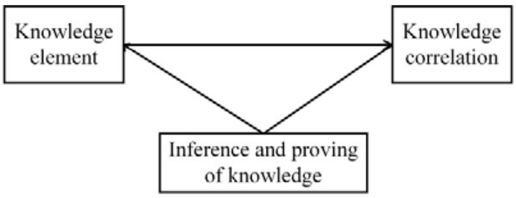


Fig. 2. Relations between inference and proving of knowledge, knowledge correlation and knowledge element

1) Knowledge element. Knowledge element refers to a knowledge concept in some domain knowledge structure (including basic definition, theorem and axiom etc.). Every knowledge element is subject to set K for description:

$$K = \{a_i \,|a_i \text{ is a attribute of knowledge element}\} \,. \tag{1}$$

Elements in sets express basic characteristics of the knowledge element. Extension of knowledge element is instance. In order to describe evaluation for every instance to attribute of knowledge element, one set V_{a_i} is used for expression for range of every attribute. The whole range V of concept for knowledge element:

$$V_{a_i}\!=\{v_j \,|v_j \text{is a value of attribute } a_i\} \,. \tag{2}$$

In (2) form, $V = \bigcup\limits_{a_i \in A} V_{a_i}$

2) Knowledge correlation. Different knowledge elements need to be defined or proved in a completed domain knowledge structure. Some of those knowledge elements need to be defined firstly, but others are based on knowledge element defined previously. One set C can be used to describe various knowledge elements in knowledge structure:

$$C = \{A_i \,|A_i \text{ is a knowledge element in knowledge structure}\} \,. \tag{3}$$

There may be correlation between knowledge elements in knowledge structure, and there may be no correlations. In general, correlations of knowledge element are subject to the following circumstances (providing that there are knowledge elements x,y, z in knowledge structure):

①$x < y$: shows knowledge element x is preface knowledge of y, and y is subsequent knowledge of x. ② in case $x < y$ and $y < z, x < z$: it shows that in case x is preface knowledge of y and y is preface knowledge of$z$, x is preface knowledge of$z$. ③ in C, $x < x$is false, then that knowledge element x is its own preface knowledge is not allowed to appear. ④ in case x is not preface knowledge of$y$ and y is also not preface knowledge of x, and x and y is mutually independent.

3) Inference and proving of knowledge. Knowledge in one knowledge structure shall be correct and consistent. At the same time, completion of one knowledge structure is relative, and it shall continuously improve itself with the increase and upgrading of knowledge. In order to guarantee the above characteristics of knowledge structure, mechanism for inference and proving of knowledge shall be established. Then it can be subject to logical form to judge whether relations of knowledge elements in domain knowledge structure and relations between knowledge elements are correct and consistent or not, and self-improvement of knowledge structure shall be completed. Correctness and consistency of and between knowledge elements and self-improvement of knowledge structure are related. Initial relations of and between knowledge elements are planned according to prior knowledge of human. They are extracted from information acquired by human, and its treatment to these pieces of information shall be guaranteed correct. But when information continuously increases and changes, new knowledge needs to be summarized. At the same time, it can affect original knowledge, which can cause changes for correctness and consistency of knowledge. Inference and proving of knowledge is an important support for assuring completion of one domain knowledge structure.

3. Construction for self-organizing neural network of graph of domain knowledge ontology

3.1. Model structure for self-organizing neural network

Model of self-organizing neural network is structure model for multi-layer tree network [4] made up of input layers and competition layers (that is output layer). Every input node and all nodes in neural tree of this model are connected through weight W to realize nonlinear descending dimensional mapping of input signal. When it arrives at node of the same tree, input mapping shall be kept with topology invariance. Neuron quantity in input layer is line number or row number of fuzzy similar matrix (that is quantity of samples in this set), and it is shown in Fig.3. Sorting results can be expressed in competitive layer through repetitive learning of input, capturing mode characteristics included in all input modes and providing self-organization to it in such structure. When network accepts input similar to mode remembered, network will recall this kind of mode for correct sorting. For mode that does not exist in network memory, under premise of not affecting existing memory, self-organization neural network will memorize this new mode.

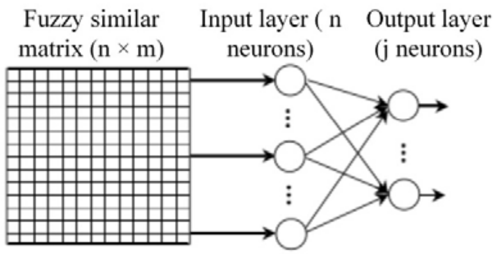


Fig. 3. Structure model for self-organization neural network

Model learning sample is made up of sample having N sorting indexes. Providing that obvious category of these points of N-dimension space is same or samples similar in some features is relatively close in N-dimension space, samples that are close constitute one category. They form a cluster in N-dimension space. When input sample belongs to several types, N-dimension space will present characteristic of multiple clustered distribution. Every cluster represents one type, and cluster center is clustering center of type. Distance between samples of the type and clustering center of this type shall be small. This distance can be measured at the usage of Euclidean distance:

$$D_j = \sqrt{\sum_{i=1}^{N} (x_i - W_{ij})^2}\,. \tag{4}$$

Where, x_i is sorting indicator; W_{ij} is clustering center for dynamic type the j type; D_j is Euclidean distance.

Algorithm for self-organization neural learning does not need teaching signal, and it judges sample type according to Euclidean distance of sample and clustering center, and algorithm step is shown as follows:

Step 1: providing threshold value β, β is used to control thickness of sorting. The larger the β is, the thicker the sorting is. The lesser the type quantity is; the smaller the β is, and the sorting is finer. Type number is more. Therefore, confirmation of β value shall be trialed according to specific conditions.

Step 2: providing that quantity for initial neuron of output layer is 1 (that is j=1), one learning sample shall be chosen for taking connection weight W_{ij} as the initial value.

Step 3: input one new learning sample, and calculate its Euclidean distance D_j with clustering center W_{ij} of every dynamic type.

Step 4: output neuron having the smallest Euclidean distance D wins through competition:

$$D_j^* = \min\{D_j\}\,. \tag{5}$$

Step 5: in case $D_j^* < \beta$, it is deemed that current output sample belongs to dynamic type for representatives of output neuron, and connection weight W_{ij} is adjusted as follows:

$$W'_{ij} = (x_i - W_{ij})/h_j\,. \tag{6}$$

Where, W'_{ij} is the adjustment value ofW_{ij}; h_j is the quantity for current samples of the j type dynamic. Then it can be transferred to step 3).

Step 6: in case $D_j^* \geq \beta$, it shows that this output neuron wins through competition. But current input sample can not be deemed as belonging of dynamic type for representative of this output neuron, and it belongs to new type. Therefore, one $j = j+1$ shall be added to output neuron to express new dynamic type. This output sample is taken as initial value of$W_{i(j+1)}$. Then it can be transferred to step 3.

Step 7: such circulation shall be conducted until all samples are learned. Ultimate quantity for output neuron of network model is quantity for types of all samples, and connection weight is clustering center value of all dynamic types.

The above learning algorithm shows that self-organization neural network has characteristic of plasticity and self-organization. At the same time, the process of learning and training of network is the process of dynamic sorting for measure data. After training is completed, established network model is the sorting model. At the time of acquisition of new measured data, network model can be input, and dynamic type for representative of output layer neuron that wins finally through competition is type that this sample belongs to, which is dynamic identification process of model to new data.

3.2. Frame of ontology learning (KLSeeker)

How to generate ontology graph (DOG) is described in this section and classification of domain knowledge (DocOG) based on domain ontology graph can be realized. KLSeeker is a completed system frame, and it defines and realizes four components; research contents are shown in the following: (1) modeling of ontology graph (ontology graph structure); (2) ontology learning (learning algorithm); (3) ontology generation (generation process); (4) ontology searching (operation for information searching system). KLSeeker system can be used to develop various modules for intelligent applications based on ontology which is subject to four definitions. Therefore, the whole KLSeeker system frame is divided into four modules for disposing process for ontology of different kinds, and it is shown in Fig.4.

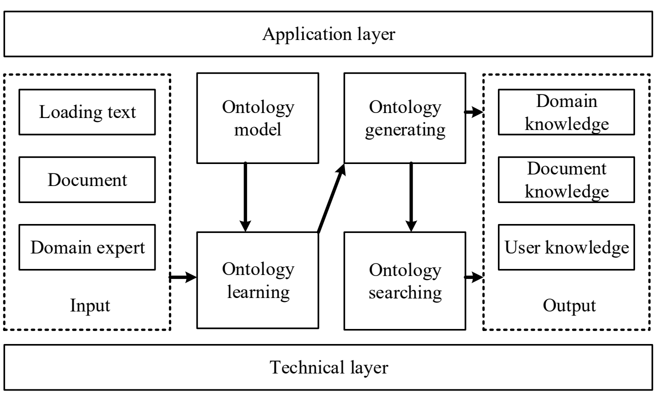


Fig. 4. Basic frame for klseeker system

Ontology graph is a new method used in KLSeeker system model for establishing of domain knowledge ontology. Ontology graph is made up of concept unit of different layers, and it is subject to different types of relations for correlation. Its essence is one vocabulary system, and it expresses concept set through mutual relations. Network model is formed through different concept units.

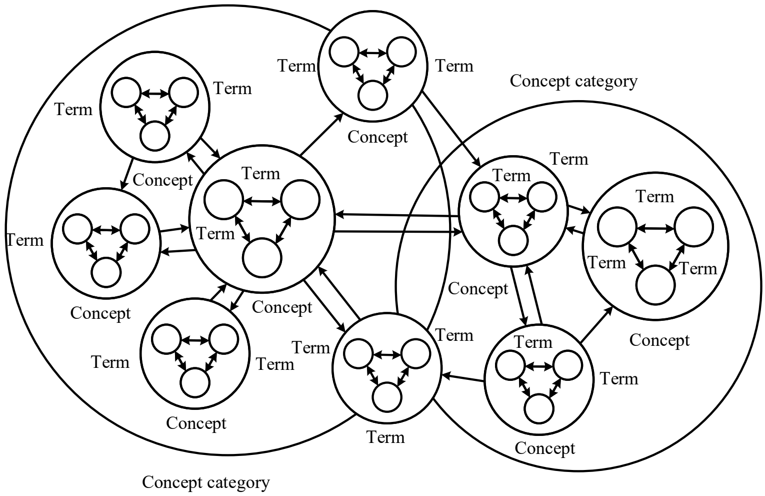


Fig. 5. Concept view

It is further shown in Fig.5 that establishing and learning process for concept view and algorithm of ontology graph model based on nodes and relation structure. According to complexity degree of its knowledge representation, ontology graph can be divided into concept unit of four types (CU). It is defined herein that ontology graph of four customers of any object and its domain knowledge representation. All customers contact through related concepts, including region (domain) knowledge representation for the whole concept structure and model of ontology graph.

3.3. Construction for graph of domain ontology

In KLSeeker, one group of concept model for ontology graph shall be defined of which concept is constructed by one group of term and its mutual relations. Firstly, definition of DOG is shown as follows:

$$OG_d = (T, F, H, R, C, A) \ . \tag{7}$$

Definition 1: (DOG) in system of KLSeeker, DOG is ontology graph related to specific domain, it can be defined as:

$$OG_d = (T, F, H, R, C, A) \ . \tag{8}$$

Herein , d defines correlation domain of ontology graph; T is set of term t_i in OG_d; H is set of sorting relation of T; R is relation set between t_i and t_j, $t_i, t_j \in T$;

C is set of clustering $t_1, \cdots, t_n$, $t_1, \cdots, t_n \in T$; A is relation feature of R.

During the process of ontology learning, DOG is acquired through sorting of domain knowledge base. Manual process includes: list for definition initial term (it can be obtained from existing dictionary), word function of definition and mapping category (it also can be obtained from this dictionary), and concept group shall be tabbed. Automatic process includes: extraction of domain terms, extraction of term relations and extraction of concept clustering. Algorithm 1 gives the generating procedure of DOG.

Algorithm 1 An variant of modified non-dominated ranking system.

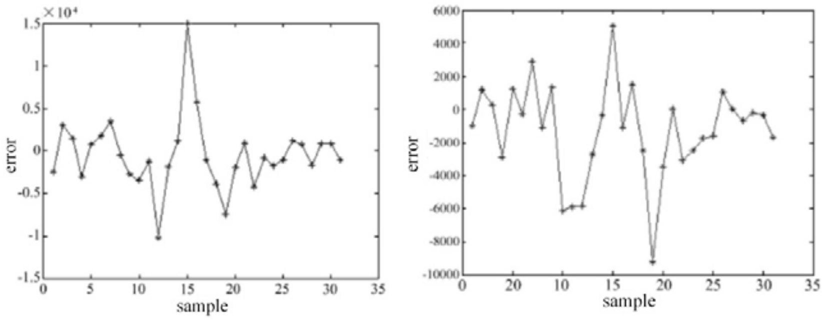
Input: T, F, H, R, C, A
Output: model for ontology graph concept of DOG
Obtain term list
for $t_i \in T$ **do**
  Generate node n_i in DOG;
  Conduct node tabbing according to term name;
  Conduct node weight evaluation according to measurement $\chi^2_{t,d}$ of term category;
end for
Obtain dependent measure of term list T, and make θ as the minimum measurement value in DOG
for every term dependent mapping, **do**
  if measures value of $t_a, t_b = \theta$ **then**
    Generate edge e_i of t_a, t_b in DOG;
    Set e_i weight as $\chi^2_{t_a,t_b}$;
  end if
end for
Remove all unconnected weights in DOG

4. Experiment analysis

4.1. Performance experiment for self-organization neural network

Hardware parameters: processor i7-6800HQ, internal storage 6G ddr3-1600, system win7 flagship version. Contrast algorithm shall be subject to BP neural network algorithm. Samples of test set shall be input into model for self-organization neural network and model for BP neural network respectively to conduct simulation test, and error sum, root-mean-square error and error percentage of two model tests output are shown in Fig.4.

It can be known according to Fig.6 data that performance for algorithm of self-organization neural network increases about 25% compared with algorithm of BP neural network in error sum index, which represents performance advantages of self-organization neural network in evaluation. Extracted algorithm of self-organization



(a) Prediction error of BP neural network (b) Prediction error of self-organization neural network

Fig. 6. Comparison of prediction error

neural network is better than that of BP neural network in fitting degree of true expectation curve, and it has lower prediction error.

4.2. Experiment data set

Test sets of different sources and properties shall be chosen: test set for Chinese website (PKU), test set for Sogou classification and test set for Fudan University classification. Information of test set is shown in Table 1.

Table 1. Specific information of data set

Data set	PKU	Sogou	FD
Quantity of domain knowledge	15511	17910	19636
Category quantity	12	9	20
Length (bits)	6-307152	118-194195	78-67218
Length mean value / variance	4813-67956324	2816-18563254	1007-54652589
Scope for quantity of domain knowledge	139-2765	1986-1986	51-3218
Category domain knowledge Quantity mean/ variance	1287-865423	1986-0	983-1156284

In Table 4, value for length scope of domain knowledge of PKU test set is relatively large, and value for length scope of domain knowledge of Sogou test set is relatively small. Mean for length of domain knowledge of FD test set is the largest. Change scope for quantity of domain knowledge in every category of FD test set is the largest, but mean quantity for domain knowledge included in its category is the minimum. Quantity of domain knowledge included in category for Sogou test set is equal which is 1990.

4.3. Comparison and analysis of performance

It shall be subject to contrast algorithm for choosing literature [13-15], of which corpus for large-scale domain knowledge shall be provided with non-negative tensor factorization by literature [13] to realize knowledge matching of ontology concept domain; literature [14] shall be provided with design for sorting technology of domain knowledge based on data set of ontology graph. These two kinds of algorithm are oriented at non-domain knowledge for algorithm design. Based on sorting method for Chinese domain knowledge of apparent domain knowledge and potentiality, literature [15] is a sorting algorithm for domain knowledge specially oriented at Chinese. For threshold value $\theta = 0$, experiment comparison result is shown in Table 2.

Table 2. Comparison result of experiment

Test set	Indexes	Literature [13]	Literature [14]	Literature [15]	ODSDC
PKU	Accuracy	86.3%	88.5%	93.4%	96.7%
	Recall rate	75.6%	81.4%	91.5%	95.4%
	F Measurement value	80.6%	84.9%	92.4%	96.0%
Sogou	Accuracy	84.3%	89.7%	92.5%	95.8%
	Recall rate	85.2%	83.4%	93.6%	97.8%
	F Measurement value	84.7%	86.4%	93.0%	96.8%
FD	Accuracy	89.6%	93.4%	95.9%	98.1%
	Recall rate	82.3%	85.4%	90.5%	92.3%
	F Measurement value	85.8%	889.2%	93.1%	95.1%

It can be known from comparison result for test of Table 5 that classification accuracy, recall rate and F measurement index for Chinese domain knowledge of ODSDC algorithm are all better than contrast algorithm compared with literature [13-15]. For algorithms designed in literature [13-14] are oriented at domain knowledge and ontology graph, effect of its application in classification of Chinese domain knowledge is not ideal. Literature [15] is algorithm design aimed at Chinese corpus, therefore its algorithm effect is better than algorithms of literature [13-14] and worse than ODSDC algorithm.

5. Conclusion

A construction method for spectrum of domain knowledge ontology based on concept study of self-organization neural network is presented in this Text. According to structure model for domain ontology graph, frame of ontology learning shall be constructed. It shall be subject to model for self-organization neural network for construction of fuzzy clustering method to realize sorting algorithm for KLSeeker knowledge ontology of dichotomy. Algorithm effectiveness is verified in experiment results. Therefore, for ontology graph is simplification of application, and conceptual relationship type can not be provided with ontology study through semi-supervision mode. In addition, effective ontology verification is not enough to measure completion of generating DOG.

References

[1] Shi S, Cao J, Li F, et al.: (2014) *Construction of a technique plan repository and evaluation system based on AHP group decision-making for emergency treatment and disposal in chemical pollution accidents*[J]. Journal of Hazardous Materials, 276(9):200-206.

[2] Jian M, Fan Z, Wei Q: (2007) *Existence and construction of weight-set for satisfying preference orders of alternatives based on additive multi-attribute value model*[J]. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 31(1):66-72.

[3] Jian M, Fan Z, Wei Q: (2007) *Existence and construction of weight-set for satisfying preference orders of alternatives based on additive multi-attribute value model*[J]. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 31(1):66-72.

[4] Jiménez A, RíOs-Insua S, Mateos A:1 (2003) *A decision support system for multiattribute utility evaluation based on imprecise assignments*[J]. Decision Support Systems, 36(1):65-79.

[5] Cunbin L I, Gao P: (2015) *Risk Group Evaluation Based on Cloud Model and Grey Incidence Decision-Making for Important Power Consumers*[J]. Electric Power.

[6] Liu B, Huo T, Liao P, et al.: (2015) *A Group Decision-Making Aggregation Model for Contractor Selection in Large Scale Construction Projects Based on Two-Stage Partial Least Squares (PLS) Path Modeling*[J]. Group Decision & Negotiation, 24(5): 855-883.

[7] Wu Y, Lin P, Chen W: (2011) *Credit evaluation of construction-agency based on entropy AHP multi-attributes improved TOPSIS decision model*[C]// International Conference on E -Business and E -Government. IEEE, 2011:1-5.

[8] Yao Y, Kou Z, Meng W, et al.: (2014) *Overall performance evaluation of tubular scraper conveyors using a TOPSIS-based multiattribute decision-making method.*[J]. The Scientific World Journal, 2014, (2014-5-29), 2014(6):753080.

[9] Yang J, Zhang Y N, Business S O, et al.: (2014) *Project Transaction Mode Selecting Based on Group Decision-making of Fuzzy Multi-attribute*[J]. Journal of Civil Engineering & Management.

[10] Zhang Y, Wei Q, Li Y, et al.: (2012) *Construction of Significant Hazard Sources Evaluation Model Based on Multiple Attribute Decision Making with Intervals*[J]. Advances in Information Sciences & Service Sciences, 4(15):404-410.

[11] Liu G F, Wang L, Zhou Z, et al.: (2016) *Optimization of Slope Treatment Schemes Based on Gray Fuzzy Multi-attribute Group Decision Making*[J]. Journal of Yangtze River Scientific Research Institute.

[12] Wei S U, Wang J: (2012) *The Multi-Attribute Evaluation Method Based on Maximize the Variance for Building Construction Bidding*[J]. Mathematics in Practice & Theory, 42(12):17-20.

[13] Huang W, Lei Z: (2008) *Credit Evaluation of Construction Enterprises Based on Fuzzy Multi-attribute Decision Method*[J]. Construction Economy.

[14] Rouyendegh B D: (2012). *Evaluating Projects Based on Intuitionistic Fuzzy Group Decision Making*[J]. Journal of Applied Mathematics,2012,(2012-7-3), 2012(1):203-222.

[15] Zhang, Z., Ou, J., Li, D., & Zhang, S.: (2017). *Optimization Design of Coupling Beam Metal Damper in Shear Wall Structures.* Applied Sciences, 7(2), 137.

[16] Weisen Pan, Shizhan Chen, Zhiyong Feng.: *Automatic Clustering of Social Tag using Community Detection.* Applied Mathematics & Information Sciences, 2013, 7(2): 675-681.

[17] Cui Liu, Yanhu Li, Yingyue Zhang, Chuluo Yang, Hongbin Wu, Jingui Qin, and Yong Cao: *Solution-Processed, Undoped, Deep-Blue Organic Light-Emitting Diodes Based on Starburst Oligofluorenes with a Planar Triphenylamine Core, Chemistry* – A European Journal, 2012, 18(22), 6928-6934.

Received May 7, 2017