Network-based relevance relationship generating for empirical engineering knowledge

Xinyu Li[1], Zuhua Jiang[1], Bo Song[1], and Lijun Liu[1]
[1]Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai 200240, PR China

Abstract - *Cognizing and utilizing the relevance relationship of knowledge are evitable issues for enterprises and organizations to maintain preponderance. However, the grueling analysis of relevance relationships, especially for empirical knowledge in engineering field, has been manually processed by domain experts. In this paper, an automatic network-based relevance relationship generating method is proposed for representing the relations among empirical engineering knowledge (EEK) and assisting in comprehending structure of the engineering domain. Two phases, EEK elicitation and formalization as well as EEK networks foundation, are included in the generating method and implemented with natural language process, sematic similarity calculation and fuzzy neutral network prediction techniques. Relevance network of empirical knowledge in computer-aided design (CAD) is constructed and verified by domain experts and long-term practitioners. Experimental results show that the proposed method outperforms the former approaches in feasibility and effectiveness, and thereby offer a way for further understanding the evolution course of EEK.*

Keywords: Empirical engineering knowledge; network; relevance relationship; knowledge representation; data visualization

1. Introduction

In the era of knowledge-driven economy, emerging knowledge based on the long-existing concepts, techniques, methodologies, experiences and activities, as well as the best management and utilization of it, is the key to maintain the competitiveness preponderance of the organizations and enterprises in creativity and adaptability [1]. And with the development of the Internet and information technology, knowledge is presented with rapid transmission and multiple interdisciplinary, which further promotes the complexity of the relevance relationships among knowledge [2]. How one piece of knowledge links to others and what potential information is concealed behind that relationships are two urgently answered and increasingly evitable problems in the field of knowledge management. The answers of the two problems could help the intellectual workers to find a breakthrough for the facing questions, obtain a comprehensive cognition for the engineering field, and grasp the direction of the future research. However, current analysis of the relevance relationship of knowledge mainly depends on the domain experts, which may be difficult to be comprehensive and objective because of the intrinsic time-delay and capacity-limitation of the experts. Such defects are more critical in feasibility and effectiveness of the expert-relied analysis when analyzing a proliferating field with huge amount of information and knowledge. Therefore, the adaptation of an automatic and scientific method for filtering colossal information and mining persuasive relationships should become a task of top priority.

New ideas and concepts are often the consequences of the original ones, which is the fundamental relevance mechanism in knowledge development [3]. In the field of academia, the researchers used metrology method to collect and census the keywords, abstracts and cited references, and, hence, speculated the relevance relations [4-7]. However, the method of metrology may not achieve a satisfied result when it is deployed into the field of engineering, especially faced with the empirical engineering knowledge (EEK). Two main reasons lead to this unavailability: (1) Although there is some clear-coded and correct-recognized knowledge in engineering field, such as the formulas, standards and specifications, most of the engineering knowledge is derived from the solutions of the actual engineering missions and presented as non-canonical knowledge with scenario dependency, concepts ambiguity and correctness uncertainty [8, 9]; (2) The precise expression of the subject knowledge via several explicit concepts and propositional logics is the fundamental of the literature statistic and quantitative analysis, while the concepts and logic relations in EEK is concealed in the records formed by natural language and incapable to elicit directly [10].

Some scholars dealt with relationships with the network methods. Pyka A. et al. [11] simulated the innovation in modern knowledge-based industries by an agent-based network model. Liu J. et al. [12] presented two knowledge-generation models via the hyper-network and analyze the distribution of knowledge stock, which could be helpful for deeply understanding the scientific research cooperation. Lee K.M. et al. [13] expressed the knowledge in Bayesian networks and proposed an agent framework. Although such network-based methods put forward some new ideas for detecting knowledge relationship, only simulation models were established in their works, rather than the detailed analysis of the specific field with large amount of knowledge, which led to a questionable feasibility.

Considering the advantages and shortages of all above research works, this paper proposes an automatic two-phase network-based relevance relationship generating method combined with natural language process, sematic similarity calculation, and fuzzy neutral network prediction based on the

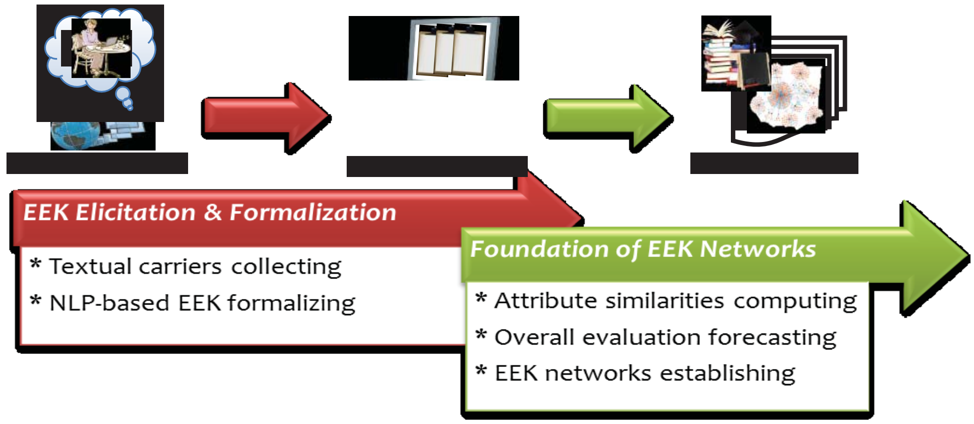

Fig. 1 Framework of network-based relevance relationship generating method

elicitation and formalization of EEKs. With few human interventions in annotating the samples, the proposed method performs well in discovering the numerical and semantic relevance relationships behind huge amount of existing actual EEKs. The feasibility and advancement of the proposed method are testified with the typical EEKs of computer-aided design (CAD) from 2011 to 2015, and verified by some domain experts and long-term practitioners.

The remainder of this paper is organized as follows. Section 2 designs the general framework of the proposed generating method. The elicitation and formalization of EEKs is presented in section 3. Section 4 details the foundation of EEK networks with calculation of attribute similarities and fuzzy neutral network prediction. The example of using the proposed method to generate the relevance network of EEKs originated from accomplishing of engineering design mission using AutoCAD software is presented in section 5. Last section includes the comparison with former works and concludes the paper with some possible improvements.

2. Framework of proposed method

Oriented to generate the relevance relationships of empirical engineering knowledge (EEK) automatically, this paper proposed a two-phase generating method based on natural language process (NLP), sematic similarity calculation, and fuzzy neutral network (FNN) prediction. Figure 1 presents the framework of proposed method.

(1) Eliciting and formalizing EEKs: Collected from threads in professional virtual communities, meeting notes, email exchanges, success or failure cases, revision history of a Wiki page and other electronic documents, available textual carriers of EEK are gathered and analyzed. Seven attributes of EEK, namely Engineering Problem, Problem Context, Problem Solution, Feature Association, Effectiveness, Contributor and Time, are extracted with part-of-speech tagging, sentences parsing, word weights computing and other natural languages process (NLP) techniques, and hence the form of EEK is constructed with $EEK = < EP, PC, PS, FA, E, C, T >$.

(2) Founding EEK networks: Similarity of each pair of attributes in two EEKs is calculated with their numerical relationships and semantic relationships. Based on seven attributes similarities, an overall evaluation of EEK relationship is forecasted by T-S Fuzzy Neutral Networks (T-S FNN). The network of EEK is founded with the EEK pairs whose strength of relationships over a threshold, and EEK networks are saved with undirected weighted graphs, and visualized by data visualization software.

3. Elicitation and formalization of EEKs

3.1 Definition and carrier of EEKs

In the enterprises, engineering experiences are of significant value for innovative design and decision-making process, possessing an indispensable part of the corporate knowledge base. Many scholars devoted themselves to the definition of empirical engineering knowledge, as well as the subsequent empirical knowledge acquisition and reuse [8, 9, 14-17]. Concluding from the related research works, the empirical engineering knowledge (EEK) processed in this paper is defined as *a consequence of probable association and extension of engineering concepts and engineering objects under specific constraints of engineering scenarios, obtained through repeated observation and practice of engineering technicians in long-term engineering activities*. Composed in personalized and ambiguous natural language, the specific engineering problem, the problem context, the solution of the problem, and the feature association with other EEKs induced from different scenarios, are described in an EEK.

In widely adopted collaborative network working environment, increasing number of EEKs are recorded and spread with the form of electronic documents existing inside and outside the enterprises, such as virtual community Q&As, meeting notes, success or failure cases, revision history of a Wiki page or other electronic documents. In these documents, the major ingredients are the textual carriers of a specific EEK happened in an actual engineering mission. Several documents also use images, videos, audios, program files and mathematical models for implementation, which is not considered in this paper. Figure 2 shows a textual EEK carrier downloaded from a professional CAD virtual community.

These textual carriers vary in the word choice and sentence building, but they all have a similar generative process. Beginning with one or several key VO structures describing the propose or the topic of the engineering problems, such as "fix layer" in the title of the carrier presented in figure 2, the question askers descript the engineering scenarios around these key VO structures, and will receive growingly admissive solutions in the interactions with the respondents. When the last interaction is completed, the EEK is generated with all its attributes recorded.

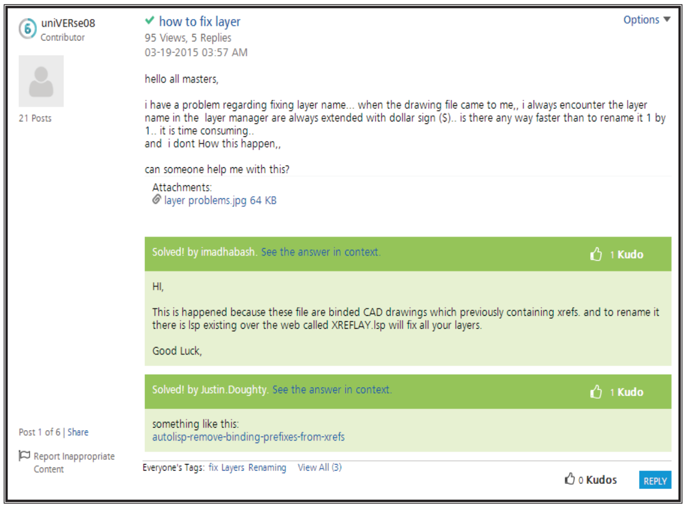

Fig.2 An example of textual EEK carrier

3.2 NLP-based EEK formalization

Certainly, the generation of textual EEK carriers is not always as perfect as the above process described. Since the askers and respondents are often unable to grasp the key to the engineering problems, there may be plenty of “noise” formed from the unrelated concepts existing in the discussion, and the same concepts may be stated in different forms by diverse participants. What’s more, for a textual record that contains no succinct topic, for example, the non-topic meeting discussion or Wiki pages revision, the key VO structures are not directly given and should be summarized from the comprehensive understanding of the whole context, which all bring difficulties for computer automated processing.. Therefore, the textual carriers of EEK should be processed with natural language processing (NLP) techniques and generated in appropriate structures before they are used for relevance relationship generation.

According to the definition in section 3.1, this paper uses seven corresponding attributes to structure an EEK, namely *EEK* = < *EP, PC, PS, FA, E, C, T* >. The content of each element is listed as follows:

(1) *EP* (Engineering Problem) proposes a specific engineering problem.

(2) *PC* (Problem Context) descripts the background informations and constrains of this *EP*.

(3) *PS* (Problem Solution) shows the empirical solution.

(4) *FA* (Feature Association) lists the relationships and corresponding strengths between other *EEK*s.

(5) *E* (Effectiveness) evaluates the fitness of *PS* in this *EP* under such *PC*.

(6) *C* (Contributor) collects all the participants in the generative process of this *EEK*.

(7) *T* (Time) records the time when this *EEK* finally formed.

In addition, for each piece of *EEK*, a unique *EEKID* is generated for indexing.

In order to filter out the unrelated informations and elicit the main attributes, Song et al. [18] adopted a method that used the Conditional Random Field (CRF) text classification techniques to label the role of each sentence and elicited problem objectives and context constrains of the empirical knowledge. Shah C. et al. [19] extracted various features from questions, answers, and the posters to select best answers through a prediction model constructed by Logistic Regression. This paper combines two methods and formalizes the EEKs.

Using CRF to label the role of each sentence (*QUESTION / CONTEXT / ANSWER / PLAIN*) in the textual carriers is the first step. Then we elicit the VO structures from the sentences labelled *QUESTTION* and use them as *EP* in an *EEK*. The singularized noun-phrases in the sentences labelled *QUESTTION* or *CONTEXT* will form *PC* in this *EEK*, and all the noun-phrases are converted into lowercase and repeated ones are eliminated. For the *ANSWER* sentences, we organize them with their original posters and evaluate them using the feature-based Logistic Regression Model. The highest regression value is chosen to be *E* of *EEK* and the set of singularized noun-phrases in corresponding *ANSWER* sentences will be *PS*. *C* and *T* of *EEK* can be obtained directly from the textual carriers, while *FA* is left blank at present and will be filled in when calculating the EEKs attribute similarities. An example of structured EEK elicited from figure 2 is shown in table 1.

Table 1 A formalized EEK elicited from figure 2

EEKID = 2494
EP = { fix layer; fix layer name } (**Noun-Phrase**: { layer, layer name});
PC = { layer name, drawing file, layer manager, dollar sign};
PS = { file, cad drawing, xref, lsp, web, xreflay.lsp, layer};
FA = { };
E = 0.645;
C = { universe08, imadhabash, beekeecz, justindoughty };
T = 2015.302

4. Foundation of the EEK networks

4.1 Similarity calculation for EEK attributes

The evaluation of the relevance relationships between a pair of EEKs is the basis of establishing EEK networks. Due to the formalization of EEKs, EEK could be expressed with seven attributes. Therefore, the evaluation will be finished with similarities of each pair of attributes. For two EEKs, $EEK_1 = < EP_1, PC_1, PS_1, FA_1, E_1, C_1, T_1 >$ and $EEK_2 = < EP_2, PC_2, PS_2, FA_2, E_2, C_2, T_2 >$, seven similarities, *EPSim, PCSim, PSSim, FASim, ESim, CSim* and *TSim*, will be calculated.

On the basis of article [18], seven similarities are calculated respectively as follows:

$$EPSim = \begin{cases} 1 & EP.I \\ 0 & EP.II \\ \frac{1}{2}\left(\frac{\sum_{NP_i \in EP_1} \max_{\forall NP_j \in EP_2} NPSim\left(NP_i, NP_j\right)}{Count\left(NP_i\right)} + \frac{\sum_{NP_j \in EP_2} \max_{\forall NP_i \in EP_1} NPSim\left(NP_i, NP_j\right)}{Count\left(NP_j\right)} \right) & EP.III \end{cases} \quad \text{.........} (1)$$

EP.I When EP_1 and EP_2 have at least one same VO;
EP.II When EP_1 or EP_2 is empty;
EP.III When EP_1 and EP_2 don't have any same VO, the similarity between EP_1 and EP_2 is calculated by all the noun-phrases NP_i of EP_1 and NP_j of EP_2;

$$PCSim = \begin{cases} 0 & PC.I \\ \frac{1}{2}\left(\frac{\sum_{NP_i \in PC_1} \max_{\forall NP_j \in PC_2} NPSim\left(NP_i, NP_j\right)}{Count\left(NP_i\right)} + \frac{\sum_{NP_j \in PC_2} \max_{\forall NP_i \in PC_1} NPSim\left(NP_i, NP_j\right)}{Count\left(NP_j\right)} \right) & PC.II \end{cases} \quad \text{..........} (2)$$

PC.I When PC_1 or PC_2 is empty;
PC.II The similarity between PC_1 and PC_2 is calculated by all the noun-phrases NP_i of PC_1 and NP_j of PC_2;

$$PSSim = \begin{cases} 0 & PS.I \\ \frac{1}{2}\left(\frac{\sum_{NP_i \in PS_1} \max_{\forall NP_j \in PS_2} NPSim\left(NP_i, NP_j\right)}{Count\left(NP_i\right)} + \frac{\sum_{NP_j \in PS_2} \max_{\forall NP_i \in PC_1} NPSim\left(NP_i, NP_j\right)}{Count\left(NP_j\right)} \right) & PS.II \end{cases} \quad \text{..........} (3)$$

PS.I When PS_1 or PS_2 is empty;
PS.II The similarity between PS_1 and PS_2 is calculated by all the noun-phrases NP_i of PS_1 and NP_j of PS_2;

For FA similarities of a pair of EEKs, two assumed relationships are considered:

- Trigger Relationship: if EEK_1 will trigger EEK_2, PC_2 of EEK_2 contains EP_1 of EEK_1
- Solved-by Relationship: if EEK_1 is solved by EEK_2, PS_2 of EEK_2 contains EP_1 of EEK_1

The similarity of EP and PC or PS will measure the two relationships, thus determining the value of *FASim*, and the value of attribute FA is filled in.

$$FASim = \max\left\{ \begin{matrix} Trigger\left(EP_1, PC_2\right), Trigger\left(EP_2, PC_1\right), \\ Solve\left(EP_1, PS_2\right), Solve\left(EP_2, PS_1\right) \end{matrix} \right\} \quad \text{..........} (4)$$

$$Trigger\left(EP_1, PC_2\right) = \begin{cases} 0 & Trigger.I \\ \frac{1}{2}\left(\frac{\sum_{NP_i \in EP_1} \max_{\forall NP_j \in PC_2} NPSim\left(NP_i, NP_j\right)}{Count\left(NP_i\right)} + \frac{\sum_{NP_j \in PC_2} \max_{\forall NP_i \in EP_1} NPSim\left(NP_i, NP_j\right)}{Count\left(NP_j\right)} \right) & Trigger.II \end{cases} \quad (4.1)$$

Trigger.I When EP_1 or PC_2 is empty;
Trigger.II The Trigger-relevance is calculated by all the noun-phrases NP_i of EP_1 and NP_j of PC_2;

$$Solve\left(EP_1, PS_2\right) = \begin{cases} 0 & Solve.I \\ \frac{1}{2}\left(\frac{\sum_{NP_i \in EP_1} \max_{\forall NP_j \in PS_2} NPSim\left(NP_i, NP_j\right)}{Count\left(NP_i\right)} + \frac{\sum_{NP_j \in PS_2} \max_{\forall NP_i \in EP_1} NPSim\left(NP_i, NP_j\right)}{Count\left(NP_j\right)} \right) & Solve.II \end{cases} \quad \text{..} (4.2)$$

Solve.I When EP_1 or PS_2 is empty;
Solve.II The Solved-by-relevance is calculated by all the noun-phrases NP_i of EP_1 and NP_j of PS_2;

Function *Count*(*NP*) is the number of non-repetitive NPs in EP, PC or PS separately. And noun-phrase similarity is computed as:

$$NPSim\left(NP_1, NP_2\right) = \begin{cases} 2^L & NP.I \\ 1 & NP.II \\ \max_{Word_i \in NP_1, Word_j \in NP_2} WSim\left(Word_i, Word_j\right) & NP.III \end{cases} \quad (5)$$

NP.I When words in the corresponding positions of the two phrases are the same; L is the phrase length;
NP.II When some words of the two phrases are the same;
NP.III When all the words of the two phrases are different, the similarity between NP_1 and NP_2 is calculated by all the words $Word_i$ of NP_1 and $Word_j$ of NP_2;

Word similarity is generated with the normalized point wise mutual information:

$$WSim\left(Word_1, Word_2\right) = \log \frac{p\left(Word_1, Word_2\right)}{p\left(Word_1\right) p\left(Word_2\right)} / \log|D| \quad \text{...} (6)$$

$p(Word_1)$ is the proportion of EEKs that contain $Word_1$, and $p(Word_1, Word_2)$ is the proportion of EEKs that contain $Word_1$ and $Word_2$ simultaneously. Since the similarity of two different words should not exceed the similarity of two same words, we use the largest possible value of point mutual information log $|D|$ to normalize it.

$$ESim = \begin{cases} 0 & \text{if } E_1 = E_2 = 0 \\ \frac{\min\left(E_1, E_2\right)}{\max\left(E_1, E_2\right)} & \text{else} \end{cases} \quad \text{..........................} (7)$$

$$CSim = \begin{cases} 0 & \text{if } C_1 \cup C_2 = \varnothing \\ \frac{2Count\left(C_1 \cap C_2\right)}{Count\left(C_1\right) + Count\left(C_2\right)} & \text{else} \end{cases} \quad \text{.............} (8)$$

Function *Count*(*C*) is the number of non-repetitive contributors in contributor set *C*.

$$TSim = 1 - \frac{\left|T_1 - T_2\right|}{T_0} \quad \text{..} (9)$$

T_0 in Eq. (9) is the maximum lag of time among all possible EEK pairs.

4.2 Fuzzy evaluation of overall relationship

Seven similarities are calculated with the Eq. (1-9), and thereby available for evaluation of overall relationship of EEK pairs. A commonly used linear weighted sum method is easy to compute the overall relationship, but the result may be not cogent for two reasons. One is that the emphasis of attributes is uncertain and weight of each of seven attributes is hard to decide; the other one is that the exact numerical values will not influence the structure of the networks and the precision of the evaluation is of slight significance.

Actually, this is an overall fuzzy evaluation with multi-input and single-output, commonly appeared in expert decision-making in geological structure, management level and finance risk assessments. Some scholars used T-S Fuzzy Neutral Networks (T-S FNN) method and achieved some good results [20-23]. T-S FNN is a method that combined supervised machining-learning and fuzzy logics and its architecture is shown in figure 3.

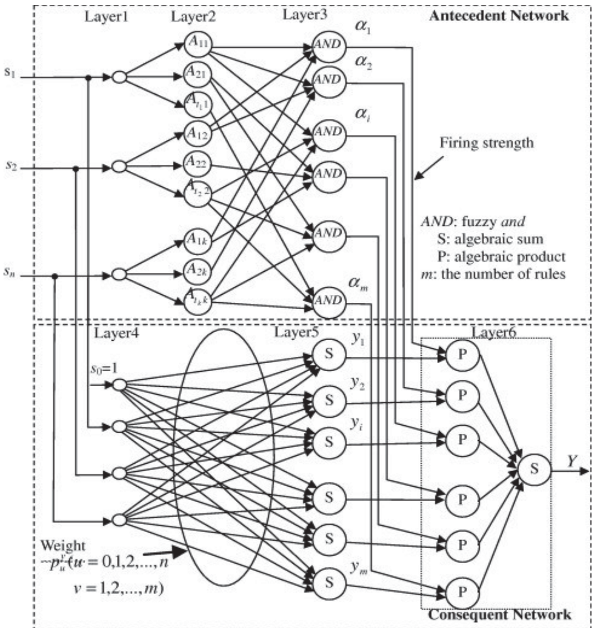

Fig.3 Architecture of the T-S FNN [23]

Layer 1 and 4 are the input layers, both receiving the values of the attributes except that a constant $s_0 = 1$ is input additionally in layer 4. Layer 2 uses Gaussian-shaped membership function $A(s) = \exp\left(-(s-c)^2 / 2\sigma^2\right)$ to fuzzificate the input attribute values, where s is the input value and c, σ are the shape parameters decided by the number of input attributes and output grades. The strength of each fuzzy rule is calculated in layer 3 and becomes the output of each neuron. Layer 5 summarizes the consequent of each fuzzy law with the function $y_j = p_0^j + \sum_{i=1}^{n} p_i^j s_i \quad (j = 1,2,\ldots$, where p_i^j is the consequence parameter and varies in the iterations. Defuzzification and normalization of all the consequents is completed in layer 6, where the overall evaluation is generated.

4.3 Establishing the EEK networks

The well-trained T-S FNN will forecast the strength of relevance relationship in any pairs of EEKs, and an undirected weighted graph $UWG = <V, E>$ representing the EEK networks is constructed in consequence. The collection of EEKs forms the vertex set V. If the relationship strength of two EEKs equals or exceeds a certain grade $Grade_{threshold}$, two corresponding vertexes are connected in the graph and the weight of the edge is their strength. The relevance network could be visualized with data visualization software for a better cognition of the structure of field.

With the relevance network, it is intuitionistic for domain experts or practitioners to find the relevance relationship among EEKs. The network will also be of significant assistance for detecting the dispersed key EEKs or concentrating on groups of several intensively related EEKs.

5. Case study

In this section, we evaluated the feasibility of proposed relevance relationships generating method. We ran the Java code on a Core i5 2.5 GHz PC with 8 GB memory, and visualized the result with Gephi software.

From three professional virtual communities, *forums.autodesk.com*, *www.cadtutor.net* and *www.cadforum.cz*, 2501 EEKs of accomplishing computer-aided engineering design missions using AutoCAD software were elicited and formalized, ranging from January 2011 to March 2015. Choosing CAD as the evaluation of proposed generating method has three reasons: (1) Computer-aided design is a typical knowledge intensive mission in the engineering field, which receives frequent attention from the engineering technicians as well as the knowledge management practitioners; (2) AutoCAD software, published by AUTODESK Corporation, is the most popular CAD tool in worldwide, and its application has been discussed deeply, widely and perennially by huge amount of CAD workers; (3) CAD experts and long-term CAD practitioners are available for analyzing the proposed process and assessing the experiment result.

We selected 320 pairs of EEKs randomly and invited 3 domain experts to scoring the pairs with a scale of 1 to 5 according to the relevance relationship in the pairs. 308 valid evaluated samples were returned and 12 samples were deleted because of the significant difference of the opinions. Actual score was determined by evaluation of the majority of experts. Table 2 lists part of the sample pairs and theirs attribute similarities.

Table 2 The attribute similarities and experts' evaluation of valid samples of EEK pairs (an illustrative part)

EEK pair		Attribute similarities							Experts' evaluation			
EEK1	EEK2	EP	PC	PS	FA	E	C	T	Rank1	Rank2	Rank3	Actual
7	586	1.00	0.30	0.29	0.54	0.94	0.00	0.97	5	5	4	5
29	1655	0.00	0.15	0.22	0.20	0.96	0.00	0.68	1	1	1	1
34	2087	0.37	0.42	0.16	0.50	0.84	0.33	0.90	2	2	3	2
58	294	0.51	0.35	0.12	0.37	0.98	0.00	0.89	4	3	3	3
66	217	0.31	0.48	0.38	0.48	0.99	0.40	0.89	3	3	3	3
71	1280	0.00	0.12	0.00	0.10	0.00	0.25	0.47	1	1	1	1
88	346	0.75	0.18	0.29	0.45	0.89	0.00	0.67	4	4	4	4

250 samples train-data and 58 samples test-data were divided and used for training of T-S FNN. In the FNN, there

were 7 input nodes, 5 output nodes and 14 fuzzy rule nodes considering the number of EEK attributes and scoring scales. Learning constant and Momentum constant both were 0.5, and the iteration number was 50000. The result of train-data and test-data is shown in figure 4.

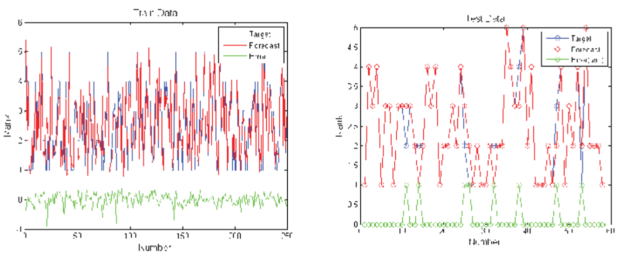

Fig. 4 The network output of training data and test data

Without the consideration of grade error, the precision of the test-data result is 86.2%, and it escalates to 100% when the admissible error is set to ±1 grade. Therefore, the relationship forecast through T-S FNN is considered reliable. The well-trained T-S FNN is utilized for evaluating any EEK pairs and forming the EEK networks. Figure 5 presents part of the undirected weighted graph of EEK networks when $Grade_{threshold}$ was set to 3. Nodes intensively located in the center illustrate the key EEKs that constructed relevance relationship with a lot of other EEKs, while ones in the margin represent the isolated EEKs with few relationships.

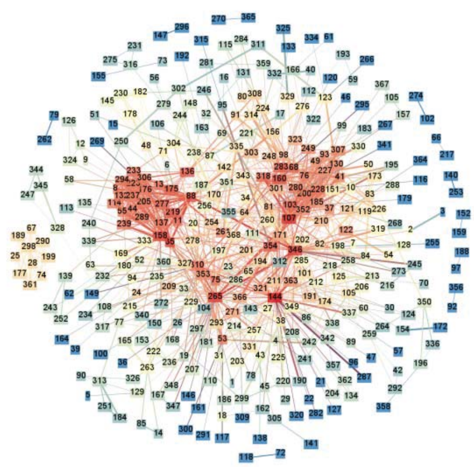

Fig.5 The undirected weighted graph of the CAD EEK relevance network (an illustrative part)

Delphi method is a useful tool for acquiring a consensus-based opinion from a panel of experts. 20 randomly selected relevance relationships in the network as well as the original EEKs textual carriers of referred EEKs were sent to 15 domain experts and long-term CAD practitioners to evaluate the validity. The performance was assessed by the questionnaire referred to Chen Y. et al. [15]. In the investigation with experts and practitioners, approximate 90% of them were satisfied with the generated relevance network in managing the empirical engineering knowledge, comprehending the domain structure of computer-aided design and mining the CAD knowledge relevance relationships with less time. Table 3 and Table 4 present the questionnaire and the result of the respondents.

Table 3 Questionnaire for assessing the performance

1. The degree to which the *CAD EEK Relevance Network* helps CAD experts and practitioners in organizing empirical CAD knowledge.
A. Very useful B. Useful C. No comment D. Useless E. Very useless
2. The degree to which the *CAD EEK Relevance Network* helps CAD experts and practitioners in understanding the structure of CAD field.
A. Very useful B. Useful C. No comment D. Useless E. Very useless
3. The degree to which the *CAD EEK Relevance Network* helps CAD experts and practitioners in mining accurate CAD knowledge relevance relationships.
A. Very useful B. Useful C. No comment D. Useless E. Very useless
4. The degree to which the *CAD EEK Relevance Network* helps CAD experts and practitioners in saving time for generating empirical CAD knowledge relationships.
A. Very useful B. Useful C. No comment D. Useless E. Very useless

6. Discussion and concluding remarks

This work developed a network-based generating method for mining relevance relationships of empirical engineering knowledge. With natural language process, sematic similarity calculation, and fuzzy neutral network prediction, the relevance network were built on the basis of the formalized elicited EEKs with seven attributes of EP, PC, PS, FA, E, C and T. The establishing of the relevance network for the field of computer-aided engineering design, as well as the assessing of the network by CAD domain experts and practitioners, has shown the feasibility and effectiveness of the proposed generating method.

Since the generating method depends on encoded empirical knowledge instead of a shared dataset, it is impossible for us to compare the proposed method with former related research work quantitatively. However, in qualitative aspects, network-based relevance relationship generating method outperforms the methods represented in articles [4-7] because of the sufficient consideration of the ambiguity and individuality of empirical engineering knowledge, and the complete combination of numerical relationships and semantic relationships among EEKs. The established relevance network using our generating method is also more persuasive than the former works. Successfully

Table 4 Respondent assessment result for the proposed modelling method

Question	Very useful	Useful	No comment	Useless	Very useless	Total	Satisfaction
Q(1)	9	5	1	0	0	15	93.3%
Q(2)	7	6	1	1	0	15	86.7%
Q(3)	6	6	2	1	0	15	80.0%
Q(4)	10	5	0	0	0	15	100.0%
Average	8.0	5.5	1.0	0.5	0	15	90.0%

operating with plenty of pragmatic empirical knowledge in actual engineering field and verified by senior experts and practitioners, the relevance network will be more cogent and obvious than the simulation models in articles [11-13].

There are several possible improvements for our methods. First, some terms used in the EEK textual carriers often mean more than they are shown literally. For example, “XREF” shown in Table 1, which is the abbreviation of “external reference”, is used as a command that allow the users to load the lines, annotations and other attachments from another drawing file in a collaborating project, which could only be matched semantically with an auto-learnt domain dictionary or a professional ontology for translating codes into meanings. Second, the proposed generating method results a static panorama of the field structure. In the future, we will try to illustrate the network dynamically along with the time and analyze the variation of the networks for the purpose of researching the evolution of empirical engineering knowledge in a long history.

Acknowledgements

The research was supported by National Natural Science Foundation of China (No. 70971085, 71271133), Shanghai Science and Technology commission (No. 13111104500), and Shanghai Municipal Education Commission (13ZZ012).

References

[1]. Dosi, G., M. Faillo and L. Marengo, Organizational capabilities, patterns of knowledge accumulation and governance structures in business firms: An introduction. ORGANIZATION STUDIES, 2008. 29(8-9): p. 1165-1185.

[2]. Žemaitis, E., Knowledge Management in Open Innovation Paradigm Context: High Tech Sector Perspective. Procedia - Social and Behavioral Sciences, 2014. 110: p. 164-173.

[3]. Palvia, P.C., S. Palvia and J.E. Whitworth, Global information technology: a meta analysis of key issues. INFORMATION & MANAGEMENT, 2002. 39(5): p. 403-414.

[4]. Yu, C., C. Davis and G.P.J. Dijkema, Understanding the Evolution of Industrial Symbiosis Research A Bibliometric and Network Analysis (1997-2012). JOURNAL OF INDUSTRIAL ECOLOGY, 2014. 18(2): p. 280-293.

[5]. Avila-Robinson, A. and K. Miyazaki, Evolutionary paths of change of emerging nanotechnological innovation systems: the case of ZnO nanostructures. SCIENTOMETRICS, 2013. 95(3): p. 829-849.

[6]. Gerdsri, N., A. Kongthon and R.S. Vatananan, Mapping the knowledge evolution and professional network in the field of technology roadmapping: a bibliometric analysis. TECHNOLOGY ANALYSIS & STRATEGIC MANAGEMENT, 2013. 25(4): p. 403-422.

[7]. Robert, C., et al., The evolution of the sleep science literature over 30 years: A bibliometric analysis. SCIENTOMETRICS, 2007. 73(2): p. 231-256.

[8]. Liu, L., Z. Jiang and B. Song, A novel two-stage method for acquiring engineering-oriented empirical tacit knowledge. International Journal of Production Research, 2014: p. 1-22.

[9]. Chen, Y., Development of a method for ontology-based empirical knowledge representation and reasoning. DECISION SUPPORT SYSTEMS, 2010. 50(1): p. 1-20.

[10]. Kump, B., et al., Tracing knowledge co-evolution in a realistic course setting: A wiki-based field experiment. COMPUTERS & EDUCATION, 2013. 69: p. 60-70.

[11]. Pyka, A., N. Gilbert and P. Ahrweiler, Simulating knowledge-generation and distribution processes in innovation collaborations and networks. CYBERNETICS AND SYSTEMS, 2007. 38(7): p. 667-693.

[12]. Liu, J., G. Yang and Z. Hu, A Knowledge Generation Model via the Hypernetwork. PLoS ONE, 2014. 9(3): p. e89746.

[13]. Lee, K.M. and K.M. Lee, Agent-based Knowledge Evolution Management and Fuzzy Rule-based Evolution Detection in Bayesian Networks. 2013 INTERNATIONAL CONFERENCE ON FUZZY THEORY AND ITS APPLICATIONS (IFUZZY 2013), 2013: p. 146-149.

[14]. Ruiz, P.P., B.K. Foguem and B. Grabot, Generating knowledge in maintenance from Experience Feedback. KNOWLEDGE-BASED SYSTEMS, 2014. 68(SI): p. 4-20.

[15]. Chen, Y. and Y. Chen, Knowledge evolution course discovery in a professional virtual community. Knowledge-Based Systems, 2012. 33: p. 1-28.

[16]. Argote, L. and E. Miron-Spektor, Organizational Learning: From Experience to Knowledge. ORGANIZATION SCIENCE, 2011. 22(5): p. 1123-1137.

[17]. Chan, F., Application of a hybrid case-based reasoning approach in electroplating industry. EXPERT SYSTEMS WITH APPLICATIONS, 2005. 29(1): p. 121-130.

[18]. Song, B., Z. Jiang and X. Li, Modeling knowledge need awareness using the problematic situations elicited from questions and answers. Knowledge-Based Systems, 2015. 75: p. 173-183.

[19]. Shah, C. and J. Pomerantz. Evaluating and predicting answer quality in community QA. in Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. 2010: ACM.

[20]. Shi, Y., H. Pan and T. Li, Evaluation model of university management informatization level based on fuzzy neural network T-S. JOURNAL OF INVESTIGATIVE MEDICINE, 2014. 62S(8): p. S108-S108.

[21]. Mosleh, M., T. Allahviranloo and M. Otadi, Evaluation of fully fuzzy regression models by fuzzy neural network. NEURAL COMPUTING & APPLICATIONS, 2012. 211: p. S105-S112.

[22]. Fukuda, S., Assessing the applicability of fuzzy neural networks for habitat preference evaluation of Japanese medaka (Oryzias latipes). 2011. p. 286-295.

[23]. Wong, W.K., X.H. Zeng and W.M.R. Au, A decision support tool for apparel coordination through integrating the knowledge-based attribute evaluation expert system and the T-S fuzzy neural network. EXPERT SYSTEMS WITH APPLICATIONS, 2009. 36(2): p. 2377-2390.