

Analysis on language fuzziness for business English in electric power industry with different development degree

RAN GUO¹

Abstract. In order to improve effectiveness of analysis on language fuzziness for business English in electric power industry with different development degree, a fuzzy meta association rule algorithm based on language fuzziness of business English to analyze hierarchy theory of evaluation model for ontology figure is proposed. First of all, evaluation model for ontology figure with language fuzziness of business English is evaluated based on analysis on language fuzziness of business English in document ontology figure and it is the process that entire KLSeeker system framework is divided into four modules to handle different kinds of ontology. Besides, with fuzzy meta association rule based on hierarchy theory, fusion algorithm of fuzzy meta association rule for language fuzziness of business English in electric power industry with different development degree has been constructed. Eventually, effectiveness of the algorithm is verified through stimulation experiment.

Key words. Electric power industry, Business English, Fuzziness, Meta association rule.

1. Introduction

With the progress of computer technology, computer technology has been introduced to business English in electric power industry and it becomes hot research. However, language fuzziness analysis on business English in electric power industry is core technology in Computer Assisted Language Learning (CALL) system. Current language learning environment and teaching model will be changed with this technology to improve language learning efficiency greatly and prompt, accurate, and objective evaluation and feedback will assist learners to find their English learn-

¹Department of Foreign Language, North China Electric Power University, Beijing, 102206, China

ing distance and their errors in English learning will be corrected during language pronounce learning[1~2].

At present, there are many representative ontology learning methods, such as literature [3~6], that are available to acquire needed ontology. However, the specific literature above needs tedious human intervention during ontology construction and learning process. These intervening measures may be in early stage. It indicates that it is during ontology extraction process before choosing match strategy with analytic hierarchy process, or at the end stage of domain ontology, namely during process of correction or reuse learning concept. Meanwhile, some scholars also consider human intervention can be put in middle stage of the algorithm for constructing concept and property of dynamic iterative learning method [7]. Because manual creation and maintenance of knowledge ontology for human is time-consuming and inefficient, simplified ontology learning method with minimum human intervention is the most practical and feasible research direction for semantic net processing and application area. What's more, it is a research with significance to learn ontology from text data because text data is an important source of human knowledge. In recent years, many ontology learning methods based on text data have been generally developed. Most researchers use methods such as machine learning and statistic analysis to develop ontology construction method for artificial intelligence and try to extract domain ontology features from text data automatically. For example, as for literature [8~10], it is more convenient and effective to study English text data with these ontology learning skills. However, due to different language features, Chinese character is more complex and diversified compared with English words so that algorithm applied to English text data does not have a good computational efficiency. Whereas, there is no successful actual practice for classification system of English text. Therefore, constructing effective ontology learning system of English text data has theoretical and practical value.

The objective evaluation for English text is mainly oriented at mastering degree of speakers for sentence rhythm and key information. Through research on intonation of English word and in combination with computer programming technology, an objective evaluation system for language fuzziness analysis of business English is designed for process of dividing entire KLSeeker system framework into four modules for treating different types of ontology. Based on of fuzzy meta association rule of hierarchy theory, that fusion algorithm of fuzzy hierarchy meta association rule for business English in electric power industry with different development degree is aimed at improving analysis effect for language fuzziness of business English.

2. Ontology figure evaluation model for language fuzziness analysis of business language

2.1. Model description

In this section, how to generate DOG will be described and evaluation of language fuzziness of business English based on document ontology figure will be realized. KLSeeker is a complete system framework, and four components are defined

and realized with it. The research content is as follows: (1) modeling of ontology figure (structure of ontology figure); (2) ontology learning (learning algorithm); (3) ontology generation (generation process); (4) ontology query (system operation of information retrieval). KLSeeker can be used for developing all kinds of intelligent application components with four defined ontology as base. Therefore, the progress that entire KLSeeker system framework is divided into four modules for treating different types of ontology is shown in Fig. 2.

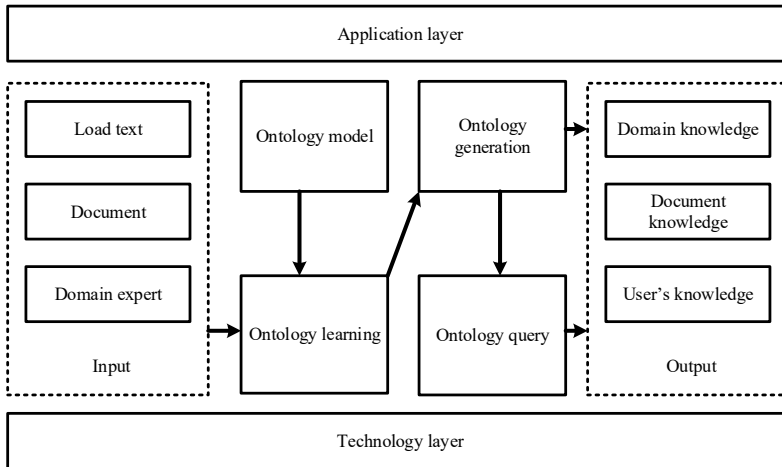


Fig. 1. Basic framework of KLSeeker system

Ontology figure is a new method of ontology creation for domain knowledge in KLSeeker system model. Ontology figure consists of different levels of concept units and it is related to different types of relations. The essence of it is- word system, and concept set is represented through mutual connection. Network model is formed with different concept units.

2.2. Word unit division

Word distribution is a factor evaluation feature of language fuzziness analysis for business English and it plays a significant role in statement organization and semantic expression. Therefore, first of all, words shall be divided for language fuzziness analysis of business English, if we need to observe layout feature of language fuzziness for business English. Division of sentences and words is mainly based on three features of language fuzziness for business English. (1) That excluding components that influence objective evaluation for sentence is included in pretreatment of word signal. (2) The feature of loudness for stressed syllable in sentence will reflect in energy intensity of time domain and it indicates word-word basic rhythmic unit presents strong word energy. According to definition for short-time energy of word

signal $s(n)$:

$$E_n = \sum_{m=-\infty}^{\infty} [s(n)\omega(n-m)]^2. \quad (1)$$

Energy value for language fuzziness of business language is extracted.

(3) Because there is certain difference of speaking speed for different people and pronunciation of the same sentence for different people will cause certain different duration of sentence. However, pronunciation of the same sentence for different people is subject to rule that unit duration of a sentence takes certain proportion of whole sentence.

(4) Due to strong word energy intensity of basic word unit, the first step for word-word basic rhythmic unit extraction can be conducted in accordance with the feature. The design is subject to double-threshold comparison method for word terminal detection. Through a large quantity of experimental verification, the following two threshold values are set in the design:

$$\begin{cases} T_u = (\max(\text{sig_in}) + \min(\text{sig_in}))/2.5 \\ T_l = (\max(\text{sig_in}) + \min(\text{sig_in}))/10 \end{cases} \quad (2)$$

(5) Because there is feature of slight long pronounce in stressed syllable of sentence, however, there may be large energy value in stressed syllable researched in first step and it indicates problem that auditory sense expression is loud pronunciation but with short duration. These units maybe short vowels that may be disturbed by signal peak and they don't constitute stressed syllable, and stressed syllable will be further sifted according to feature of light long pronounce in stressed syllable. The minimum unit of stressed syllable will be set to a approximate stressed vowel duration of 100ms. Before& after contrast of minimum unit duration is set (std indicate after set and test indicates before set): through the above steps, the division of word unit in sentence has been completed.

3. Fuzzy meta association rule based on hierarchy theory

3.1. Related definition

$I = \{i_1, i_2, \dots, i_m\}$ is designate as finite set of language fuzziness for business English in electric power industry, and fuzzy transaction is designated as non-null fuzzy subset. Item $i \in I$ in τ is satisfied with membership degree of unit interval, such as $\tau(i) \in [0, 1]$. Through extension, if i is subset of A , membership degree of fuzzy transaction item τ for A can be defined as: $\tau(A) = \min_{i \in A} \tau(i)$. Then, a group of fuzzy transaction \tilde{D} is given. Among $\tau \in \tilde{D}$, when $\tau(A) \leq \tau(B)$, fuzzy association rule $A \rightarrow B$ in \tilde{D} is established.

In hierarchy theory, item $i \in I$ and general item set $A \subset I$ are considered, and hierarchy is defined in fuzzy set \tilde{D} , where, pre-defined hierarchy set is $1 = \alpha_1 > \alpha_2 > \dots > \alpha_m > 0$. With the following functions, $\rho_A(\alpha) = \{\tau \in \tilde{D} : \tau(A) \geq \alpha\}$,

$\alpha \in \Lambda_A$ can be expressed clearly in hierarchy.

As for any item set of A and B in \tilde{D} , their hierarchy associated rule $A = (\Lambda_A, \rho_A)$ and $B = (\Lambda_B, \rho_B)$ are considered. The fourfold table is defined as $\mathcal{M}_{\alpha_i}(A, B, \tilde{D})$, and it contains cardinal number α_i of related hierarchy \tilde{D} [10]:

$$\mathcal{L}_{\alpha_i}(A, B, \tilde{D}) = \begin{bmatrix} \mathcal{L}_{\alpha_i} & B & \neg B \\ A & a_i & b_i \\ \neg A & c_i & d_i \end{bmatrix}. \quad (3)$$

Where, a_i, b_i, c_i , and d_i are nonnegative integers and there are similar definitions of $a_i = |\rho_{A \wedge B(\alpha_i)}|$, $b_i = |\rho_{A \wedge \neg B(\alpha_i)}|$, c_i , and d_i . In order to evaluate effectiveness of fuzzy association rule, the following basic probability distribution is defined [11]:

$$\text{Prob}(Y) = \sum_{\alpha_i \in \Lambda: Y = \rho(\alpha_i)} \alpha_i - \alpha_{i+1}. \quad (4)$$

Where, there must be at least one preimage $\alpha \in \Lambda$. With $\Lambda = 1$ set, clearness is conducted:

$$FSupp(A \rightarrow B) = \sum_{\alpha_i \in \Lambda_A \cup \Lambda_B} (\alpha_i - \alpha_{i+1}) \left(\frac{a_i}{a_i + b_i + c_i + d_i} \right). \quad (5)$$

$$FConf(A \rightarrow B) = \sum_{\alpha_i \in \Lambda_A \cup \Lambda_B} (\alpha_i - \alpha_{i+1}) \left(\frac{a_i}{a_i + b_i} \right). \quad (6)$$

$$FCF(A \rightarrow B) = \sum_{\alpha_i \in \Lambda_A \cup \Lambda_B} (\alpha_i - \alpha_{i+1}) \cdot \begin{cases} \frac{a_i d_i - b_i c_i}{(a_i + b_i)(b_i + d_i)}, & \text{if } a_i d_i > b_i c_i \\ 0, & \text{if } a_i d_i = b_i c_i \\ \frac{a_i d_i - b_i c_i}{(a_i + b_i)(b_i + c_i)}, & \text{if } a_i d_i < b_i c_i \end{cases} \quad (7)$$

When $|\rho_A(\alpha_i)| = 0$, there may be uncertain form 0/0 for confidence degree calculation, and this situation is forbidden. Therefore, it is set as 1 for above specified uncertainty to ensure definition of fuzzy rule.

3.2. Extraction of meta rule

Meta association rule is aimed at acquiring data position distribution and horizontal partition storage from database. Under the two situations, we hope possible association may be extracted from one association rule of each main data set. As for large-scale, complex, and heteroid language data set for business English in electric power industry, the treatment way is effective.

It is explained with an example here. Suppose there is an institution with multiple branches. For example, there are multiple branches spreading all network for language tag in business language of electric power industry. Under this situation, data stored in each separate language branch of business language will have similar structure. There will be certain privileges for extracted knowledge with meta rule: (1) it is unnecessary to treat complete data set because it will reduce calculation efficiency of algorithm. (2) Result/model is allowed to be acquired from single database to reduce needed time of rule in digging process.

Overall framework recommended by us is shown in Fig. 1 and it consists of two main steps. (1) Association rule acquired from database periodically and it can be presented as D_i ; (2) Association fusion is generated with the form of meta association rule. With the rule collected in the first step, new constructed database is named as meta database \mathcal{D} that can be expressed as fuzzy or Boole form and it is shown in Table 1. It depends on user's purpose and the accuracy of extraction rule in first step.

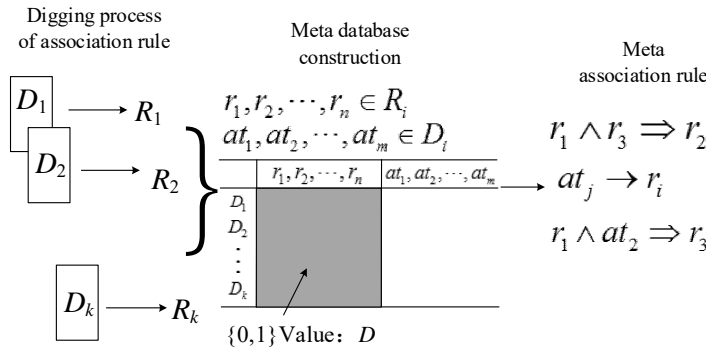


Fig. 2. From Original data set to final meta association rule

Table 1. Boole (up) and fuzzy (down)meta database

\mathcal{D}	r_1	\dots	r_n	at_1	\dots	at_m
D_1	1	\dots	0	1	\dots	1
D_2	0	\dots	0	0	\dots	1
\vdots	\vdots	\ddots	\vdots	\vdots	\ddots	\vdots
D_k	1	\dots	1	1	\dots	0

$\tilde{\mathcal{D}}$	r_1	\dots	r_n	at_1	\dots	at_m
D_1	0.2	\dots	1	0.9	\dots	1
D_2	0	\dots	0.6	0	\dots	0.2
\vdots	\vdots	\ddots	\vdots	\vdots	\ddots	\vdots
D_k	0.9	\dots	0.5	1	\dots	0.1

Then, distinguishable fuzzy or clear meta association rule is subject to the following steps:

Step 1: Let $\{D_1, D_2, \dots, D_k\}$ is a group of property sharing database. With the rule, procedure is extracted and a group of different association rule R_i can be provided for each D_i in each database. As for found association rule and its evaluation value, they can be expressed as set R_1, R_2, \dots, R_k , among others, suppose selecting the same threshold value with minimum support and certainty factor for database treatment, there are many repeated rules without losing generality.

Step 2: Information is collected from Step1 and structurized meta database \mathcal{D} has been constructed, while all rules belong to at least set R_i . The other feature of is that abundant data can be utilized to detailed describe D_i with description way of at_1, at_2, \dots, at_m . Meta database finally acts as input of meta rule extraction process.

3.3. Fusion of fuzzy hierarchy meta association rule

Digging process of meta association rule is shown in Algorithm 1. So-called item calculation process for frequent item set or candidate set is shown in code of the 16-12 line. These rules can be extracted with threshold value that over user's definition and the algorithm is shown in code of the 13-15 line. The calculation in first step is the most complex and different heuristic strategies are proposed for reducing time that spent in digging process of the rule. In proposed algorithm, items are expressed with binary string to accelerate computation speed. Besides, memory usage is not high with binary string for expression so that memory demand of system has been reduced.

In Algorithm1, in the premise of satisfying support of each meta database $\{D_1, D_2, \dots, D_k\}$ and requirement of certainty factors, digging of association rule starts. Then, meta database \mathcal{D} or fuzzy meta database $\tilde{\mathcal{D}}$ have been established. In the step, supplementary information can service as attribute addition of meta database. Attribute of fuzzy meta database $\tilde{\mathcal{D}}$ canbe modeled as fuzzy set. Eventually, clear or fuzzy meta association rule can be extracted separately. Obviously, when original data set is not available, diggingprocess of meta rule will be executed in Step 18. When fuzzy meta database is available, step 22 starts so that we can dig algorithm with any fuzzy rules. Concretely, when we execute steps 2-16, meta database \mathcal{D} services as input and acquired fuzzy meta association rule serves as output.

Therefore, fusion algorithm of meta association rule based on hierarchy theory is shown in Algorithm 2. FSupp and FCF of equations (7) and (9) are used for fuzzy evaluation in parallel level set. Especially in Step 9, as for each $\alpha \in \Lambda$, certainty factor and clearness support can be calculated independently and the calculation results of equations (7) and (9) will be weighed and calculated.

Algorithm 1: meta association rule

$D_2, \dots, D_k, at_1, at_2, \dots, at_m, minsupp, minCF;$
 $D_2, \dots, D_k, at_1, at_2, \dots, at_m, minsupp, minCF;$
 $R_1, R_2, \dots, R_k;$
Output: $R_1, R_2, \dots, R_k;$

1. **for all** D_i **do**
1. **for all** D_i
2. # D_i pretreatment
3. Read D_i , and corporate storage item I ;
4. Replace D_i to Boole database;
5. digging of #association rule
6. **if** $Supp(X) \geq minsupp$ **then**
7. $X \in C$ # C services as candidate set
8. **endif**
9. **for all** $X, Y \in C; X \cap Y = \emptyset$ **do**
10. **if** $Supp(X \rightarrow Y) \geq minsupp$ **then**
11. $X \wedge Y \in C$ and $X \rightarrow Y$ are frequent;
12. **endif**
13. **if** $CF(X \rightarrow Y) \geq minCF$ **then**
14. $X \rightarrow Y \in R_i$ and $X \rightarrow Y$ are certain;
15. **endif**
16. **endfor**
17. **endfor**
18. Creation of #meta database \mathcal{D}
19. Compile all different rules R_1, R_2, \dots, R_k ;
20. Create with compiling rule and additive property \mathcal{D} ;
21. # 21. meta rule mining
22. With \mathcal{D} , steps 2~16 are executed repeatedly;

Algorithm 2: digging of fuzzy association rule

$minsupp, minCF;$
 Input: Λ , fuzzy set \tilde{D} , $minsupp$, $minCF$;
 Output: FR , # fuzzy association rule set

1. # \tilde{D} pretreatment
2. Read D_i , and corporate storage item I ;
3. Replace \tilde{D} to p Boole database; # p serves as hierarchy number of Λ .
4. Database is coded as binary p vector.
5. digging of #association rule
6. **for all** $\alpha_i \in \Lambda$ **do**
7. Steps 6-16 in Algorithm 1 have been executed repeatedly and clear rule set is acquired in each hierarchy.
8. Read all found rules in hierarchy α_i ;
9. $FSupp$ and FCF are calculated with equation (2) and (4);
10. endfor
11. Collect rule satisfying $FSupp$ and FCF requirements;

3.4. Analysis on calculation complexity

As specified above, digging algorithm of association rule generally consists of two steps. Calculation complexity degree of the algorithm is $O(|D|2^{|I|})$. In our proposed method, because original D_i can realize parallel treatment and the first stage of algorithm 1-17 lines has same calculation complexity degree with standard algorithm, while $O(|D|2^{|I|})$ depends on transaction quantity $|D_i|$ and item quantity. In the second stage in algorithm 18-22 lines, calculation complexity degree is relevant to original database scale k , attribute quantity m , and acquired different rule quantity in the first step. Therefore, it can be concluded that its calculation complexity degree is $O(k2^{m+r})$. Because calculation process shown in Algorithm 2 is executed in parallel for $\alpha \in \Lambda$.

4. Experimental analysis

The system is aimed at improving business English level of trainer with certain English base in electric power industry with different development degree and assisting him to better handle sentence rhythm. At the time of training, language fuzziness analysis of specific business English is provided. Therefore, it can reduce influence of excessive read, less read, or wrong read of syllable on trainers. At the same time, excessive read, less read, or wrong read of syllable of trainer will lead to difference of sound duration so that marks will be deducted by the system.

Table 2. Comparison of experimental result

Student	Evaluation method	Sentence 1	Sentence 2	Sentence 3
Student 1	Algorithm in the Thesis	C	C	C
	Algorithm in literature [13]	C	B	C
	Teacher's evaluation	C	C	B
Student 2	Algorithm in the Thesis	B	C	C
	Algorithm in literature [13]	D	C	D
	Teacher's evaluation	B	C	B
Student 3	Algorithm in the Thesis	C	D	C
	Algorithm in literature [13]	B	D	B
	Teacher's evaluation	C	C	C
Student 4	Algorithm in the Thesis	C	C	D
	Algorithm in literature [13]	B	B	C
	Teacher's evaluation	C	C	D
Student 5	Algorithm in the Thesis	C	D	A
	Algorithm in literature [13]	C	D	B
	Teacher's evaluation	C	C	A

The 10 sentences recorded by authoritative English teacher of experimental record act as standard sentences, and 10 sentences recorded by 10 English students serves

as test sentences. Evaluation result consists of 4 grades, including A, B, C, and D, and hits of wonderful, good, attention, try again will be given respectively. First of all, the algorithm in the Thesis and algorithm in literature[13] will be analyzed in comparison in the experiment and difference calculation of standard sentence and mark grade will be quantified and test sentence will be conducted with divisor as A grade quantized value. At the same time, teacher's evaluation serves as measurement criteria and the result is shown in Table 2:

According to the data in Table 2, in pronunciation experiment comparison of sentence 1-3 for different students, evaluation result of acquired fuzziness analysis of business English proposed in the Thesis is more close to that of literature [13]. It shows that the method proposed in the Thesis has advantage in evaluation accuracy.

With statistic result as research object, data clustering and information fusion are processed for realizing ability evaluation of business English in electric power industry with different development degree. Test result for accuracy of evaluation and other indexes is shown in Table 3. It can be known from analysis that there is higher accuracy of teaching ability evaluation with the method in the Thesis with better use ration of teaching material.

Table 3. Comparison of performance test

Evaluation cycle	Method in the Thesis		Literature[4]		Literature[5]	
	Accuracy of evaluation /%	Use ratio/%	Accuracy of evaluation /%	Use ratio/%	Accuracy of evaluation /%	Use ratio/%
1	98.21	98.02	87.43	89.12	83.23	86.33
2	97.09	97.67	86.55	87.34	82.12	87.30
3	96.33	99.03	88.76	89.31	86.09	79.31
4	98.54	96.34	89.43	87.67	88.23	78.92

5. Conclusion

Optimized model of English ability evaluation has been studied in the Thesis, and an English ability evaluation method based on big data fuzzy meta association rule clustering and information fusion has been proposed, while analysis model of constraint parameter index for language fuzziness evaluation analysis of business English in electric power industry with different development degree has been constructed and big data information model of English ability evaluation is analyzed with recurrence quantification analysis for realizing entropy feature extraction for information with constraint feature of English ability. Clustering and integration of index parameter for English ability is realized in combination with big data information fusion and clustering algorithm of meta association rule. On this basis, corresponding teaching material distribution plan has been prepared for realizing English ability evaluation. It can be known from the research that there is higher accuracy of English ability evaluation with the method in the Thesis with improved English teaching efficiency.

References

- [1] Y. J. ZHAO, Y. H. HU, J. J. LIU: *Random Triggering-Based Sub-Nyquist Sampling System for Sparse Multiband Signal*. IEEE Transactions on Instrumentation and Measurement. 66 (2017), No. 7, 1789–1797.
- [2] N. ARUNKUMAR, K. R. KUMAR, V. VENKATARAMAN: *Automatic detection of epileptic seizures using permutation entropy, Tsallis entropy and Kolmogorov complexity*, Journal of Medical Imaging and Health Informatics, 6 (2016), No. 2, 526–531.
- [3] Y. Y. ZHANG, A. ALGBURI, N. WANG, V. KHOLODOVYCH, D. O. OH, M. CHIKINDAS, AND K. E. UHRICH: *Self-assembled Cationic Amphiphiles as Antimicrobial Peptides Mimics: Role of Hydrophobicity, Linkage Type, and Assembly State*, Nanomedicine: Nanotechnology, Biology and Medicine, 13 (2017), No. 2, 343–352.
- [4] N. ARUNKUMAR, K. R. KUMAR, V. VENKATARAMAN: *Automatic detection of epileptic seizures using new entropy measures*, Journal of Medical Imaging and Health Informatics, 6 (2016), No. 3, 724–730.
- [5] R. HAMZA, K. MUHAMMAD, N. ARUNKUMAR, G. RAMÍREZ GONZÁLEZ,: *Hash based Encryption for Keyframes of Diagnostic Hysteroscopy*, IEEE Access (2017), <https://doi.org/10.1109/ACCESS.2017.2762405>
- [6] D. S. ABDELHAMID, Y. Y. ZHANG, D. R. LEWIS, P. V. MOGHE, W. J. WELSH, AND K. E. UHRICH: *Tartaric Acid-based Amphiphilic Macromolecules with Ether Linkages Exhibit Enhanced Repression of Oxidized Low Density Lipoprotein Uptake*, Biomaterials, 53 (2015), 32–39.
- [7] N. ARUNKUMAR, K. RAMKUMAR, S. HEMA, A. NITHYA, P. PRAKASH, V. KIRTHIKA: *Fuzzy Lyapunov exponent based onset detection of the epileptic seizures*, 2013 IEEE Conference on Information and Communication Technologies, ICT 2013, (2013), art. No. 6558185, 701–706.
- [8] J. J. FAIG, A. MORETTI, L. B. JOSEPH, Y. Y. ZHANG, M. J. NOVA, K. SMITH, AND K. E. UHRICH: *Biodegradable Kojic Acid-Based Polymers: Controlled Delivery of Bioactives for Melanogenesis Inhibition*, Biomacromolecules, 18 (2017), No. 2, 363–373.
- [9] Z. LV, A. HALAWANI, S. FENG, H. LI, & S. U. RÉHMAN: *Multimodal hand and foot gesture interaction for handheld devices*. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 11 (2014), No. 1s, 10.
- [10] W. S. PAN, S. Z. CHEN, Z. Y. FENG: *Automatic Clustering of Social Tag using Community Detection*. Applied Mathematics & Information Sciences, 7 (2013), No. 2, 675–681.
- [11] Y. Y. ZHANG, E. MINTZER, AND K. E. UHRICH: *Synthesis and Characterization of PEGylated Bolaamphiphiles with Enhanced Retention in Liposomes*, Journal of Colloid and Interface Science, 482 (2016), 19–26.
- [12] N. ARUNKUMAR, K. M. MOHAMED SIRAJUDEEN: *Approximate Entropy based ayurvedic pulse diagnosis for diabetics - A case study*, TISC 2011 - Proceedings of the 3rd International Conference on Trendz in Information Sciences and Computing, (2011), art. No. 6169099, 133–135.

Received May 7, 2017

