Acta Technica 62 (2017), No. 7A, 463–474

A Study of Vague Language and Its Translation Strategies in Medical English Chinese Interpretation from the Perspective of Skopos Theory

Yang Shen[1]

Abstract. Long-distance sequencing is a main challenge for statistical machine translation. Past research work shows that pre-sequencing is one possible approach to solve the problem. In the work, we propose an pre-sequencing model based on dynamic Bayesian network by combining dynamic Bayesian network modeling with linear sequencing frame along pre-sequencing research direction. The pre-sequencing model can predict word order differences between different languages better by utilizing syntactic and semantic information extracted from abundant unlabeled data. We do experiments in machine translation task from Chinese to English and Japanese to English, and experimental result verifies effectiveness of the method.

Key words. Teleology, Medicine, English-Chinese interpretation, Fuzzy language, Translation strategy.

1. Introduction

As a kind of important prototype study method, dynamic Bayesian network conforms to cognitive process of humans. Seen from machine learning perspective, dynamic Bayesian network is a kind of extremely powerful method, and has been developed in relatively rapid speed in recent years. It is multilayer dynamic Bayesian network essentially and is applied to many fields of natural language, such as question-answering system, syntactic analysis, dependency analysis, and SMT etc[1,2]. Dynamic Bayesian network is also introduced to SMT module learning, including word alignment, language model, translation model and twisting model etc. That dynamic Bayesian network is applied to natural language processing means word embedding or representation generation actually[3]. Word embedding is low-

[1]Department of English, Xi'an Medical University, Xi'an, Shaanxi, 710021, China

dimension, dense and real-value vector, and these vectors reflect certain semanteme. Advantage of dynamic Bayesian network lies in that it changes manual extraction characteristics into automatic extraction process, which overcomes incompleteness of original manual characteristics, and can finish classification or regression task through mechanical application to a classifier or regression machine after extraction[4,5]; these characteristics are located at the same continuous space simultaneously, which is beneficial to measurement of relationship between characteristics[6].

But structural characteristics of language are not considered in modeling process of traditional dynamic Bayesian network. To solve the problem, some improved algorithms are proposed, and these improved networks conform to sequential structure of language generation and improved Bayesian network conforms to hierarchical structure of language generation[7,8]. 2 kinds of model are frequently applied to machine translation. Although SMT model in the 2 aspects makes certain progress, phrase/rule induction process is not considered truly, which causes that these dynamic Bayesian networks do not conform to translation process really. Hierarchical machine translation model has always been main model for SMT task since its proposal, it extracts initial phrase pair from bilingual corpora firstly, and then extracts hierarchical rules according to initial phrase pair extracted, and these rules have certain characteristics. Although certain semantic information is contained in syntactic structure, semantic information depicting translation process of the entire sentence is still lacked; both semantic extraction and translation process structure prediction are contained in this paper; order of Bayesian network training language shall be improved firstly, to obtain monolingual semanteme, and hierarchy of improved Bayesian network training modeling language shall be used, i.e. translation process structure prediction, and similarity function shall be used to obtain bilingual semanteme[9].

This paper starts from relatively few semantic representations in the original hierarchical modeling[10,11], use Bayesian network to obtain semantic feature representation and use hierarchical translation process in recursive modeling translation model of Bayesian network. Hierarchical machine translation model of deep recursion is mainly proposed from 3 aspects:

(1) AutoEncoder is used to make modeling to source/ target language reconstruction and translation process from source language to target language, phrase/rule semantic vector representation concerned with semanteme shall be generated in unsupervised mode, and similarity between phrase/rule vectors of source language and target language shall be taken as training objective in supervised mode. Hierarchical AutoEncoder is adopted for better modeling to initial phrase pair and rule semantic feature generated based on initial phrase. Joint training of 3 stages is made in the model, which can balance significance of each stage better.

(2) In training process, seamless connection to original hierarchical translation process shall be made, to catch structured translation process better. In decoding process, monolingual and bilingual semantic features and monolingual and bilingual sensitivity features are used to improve translation performance.

(3) Alignment information is used to guide Bayesian network training, latent variable information can be considered in semantic vector acquisition, and semantic

feature will be calculated finally according to alignment information.

2. Fuzzy language in medical English-Chinese interpretation and its translation strategy model

In this chapter, we will introduce vector representation of vocabulary learning through dynamic Bayesian network firstly; then state linear sequencing model; finally propose pre-sequencing model based on dynamic Bayesian network.

2.1. Vector representation of vocabulary based on teleological perspective

Traditional natural language processing system regards word as high-dimension sparse feature. Assumed that vocabulary of a system is V, sparse representation of vocabulary $\omega_i \in V$ is a vector with length being $|V|$, of which the i^{th} dimension is 1 and other dimension is 0. Because vocabulary of natural language is quite large, such sparse representation is a high-dimension characteristic. High-dimension vocabulary characteristic cannot describe similarity between vocabularies, natural language model making training based on it will meet data sparse problem generally, so its generalization capability is relatively weak. To improve the problem, Bengio et al. proposed language model based on dynamic Bayesian network. In their work, a dynamic Bayesian network model transforms high-dimension vocabulary characteristics into low-dimension and dense vector representation; through discriminative training on a great deal of text, the model can map vocabulary with similar context into similar point in low-dimension vector space. Along the direction, Collobert et al. proposed a kind of rapider learning method based on negative-sampling, so vector representation of a large number of vocabulary is learnt effectively in large-scale corpus.

For a text fragment $\{\omega_{-n}, \cdots, \omega_0, \cdots, \omega_n\}$ with length being $(2n+1)$, dynamic Bayesian network firstly obtains vector representation $v(\omega_i)$ of these vocabularies through a representation searching layer LOOK UP, and then splices these vectors, and obtains output of dynamic Bayesian network through a linear layer l_1, tangent hyperbolic layer *tanh*, and a linear layer l_2, which is as shown in formula (1).

$$s(\omega_{-n}, \cdots, \omega_0, \cdots, \omega_n) = l_2 \tanh \cdot LOOKUP(\omega_{-n}, \cdots, \omega_0, \cdots, \omega_n)\,. \tag{1}$$

Where:

$$l_i(x) = W_i x + b, i = 1, 2\,.$$

$$\tanh(x) = \frac{\exp(2x) - 1}{\exp(2x) + 1}\,.$$

We hope that the dynamic Bayesian network can distinguish true text fragment and false text fragment randomly generated. Specifically speaking, give a true text fragment $\{\omega_{-n}, \cdots, \omega_0, \cdots, \omega_n\}$, we obtain $\{\omega_{-n}, \cdots, \omega', \cdots, \omega_n\}$ by replacing ω_0 with ω' extracted from vocabulary randomly, and we hope that dynamic Bayesian network can give a score to fragment $\{\omega_{-n}, \cdots, \omega_0, \cdots, \omega_n\}$ that is higher than that of fragment $\{\omega_{-n}, \cdots, \omega', \cdots, \omega_n\}$ randomly generated. To realize the goal, Collobert et al. proposed optimization to following objects on a great deal of data through stochastic gradient descent method, which is as shown in formula (2).

$$L(\theta) = \sum_{allngrams} \sum_{\omega' \neq \omega_0} \max\left(0, 1 + s\left(\omega_{-n}, \cdots, \omega', \cdots, \omega_n\right)\right) - s\left(\omega_{-n}, \cdots, \omega_0, \cdots, \omega_n\right). \tag{2}$$

Mikolov et al. proposed another kind of fast learning method on vocabulary representation based on Skip-ngram. In the method, a feed-forward dynamic Bayesian network is used to establish conditional probability model of word ω and word $c(\omega)$ in its context, which is as shown in formula (3).

$$P\left(c(\omega)\,|\omega\right) = soft\max \cdot l \cdot LOOKUP(\omega)\ . \tag{3}$$

Where, l is a linear layer, input length is vocabulary representation length, and output length is vocabulary size; $soft\max$ normalizes output of l as probability. To quicken normalization, Mikolov proposed that hierarchical $soft\max$ method based on Huffman tree shall be adopted for acceleration. Stochastic gradient descent is adopted in Skipngram training, to make the maximum likelihood estimation on training data to the conditional probability. Above vocabulary representation learning method can be promoted to learning to n-gram Skipngram of vocabulary. We regard n-gram x of vocabulary as a whole and establish conditional probability model of n-gram x and word $c(\omega)$ in context (context is also word, instead of n-gram of vocabulary) through dynamic Bayesian network, which is as shown in formula (4).

$$P\left(c(x)\,|x\right) = soft\max \cdot l \cdot LOOKUP(x)\ . \tag{4}$$

Training method of n-gram representation of vocabulary is completely equal to vocabulary representation training method. Because context is also word in the model, calculated quantity of low-dimension representation of training vocabulary n-gram is equal to calculated quantity of training vocabulary representation; the difference is that n-gram quantity is greater than vocabulary, and accurate evaluation can only be made with more data. Vocabulary n-gram includes information that cannot be combined by vocabulary representation.

Through training on a great of text, vocabulary vector representation learnt by dynamic Bayesian network can map word with the similar grammar and semanteme into similar position in low-dimension space. Taking such vocabulary vector representation as characteristic and sequencing model input, we can utilize information contained in it and learn better sequencing model.

2.2. Linear sequencing model

Linear sequencing model is a kind of sequencing model. It regards sequencing problem of a sequence as sum of subproblem on pairwise element sequencing in sequence. Specifically speaking, for a sequence $\{1, 2, \cdots, n\}$ and a replacement π of the sequence, score given by linear sequencing model to the replacement is formula (5).

$$s(\pi) = \sum_{i<j} s(i, j, \pi(i), \pi(j)). \tag{5}$$

Where, $s(i, j, \pi(i), \pi(j))$ is sequencing score to (i, j), which is concerned with sequential order of (i, j) after sequencing, which is as shown in formula (6).

$$s(i, j, \pi(i), \pi(j)) = \begin{cases} s(i, j, 0), \pi(i) < \pi(j) \\ s(i, j, 1), otherwise \end{cases} \tag{6}$$

In other words, if relative sequence of (i, j) is unchanged in replacement, then their score will be $s(i, j, 0)$; if their relative sequence is reversed, then score will be $s(i, j, 1)$. Work done by Tromble and Eisner applies linear sequencing model to machine translation sequencing problem. For a source language sentence $src = \{\omega_1, \omega_2, \cdots, \omega_n\}$ needing sequencing and its one possible sequencing result $\{\omega_\pi(1), \omega_\pi(2), \cdots, \omega_\pi(n)\}$, score given by sequencing model is sum of sequencing score of each pair of word, which is as shown in formula (7).

$$s(\pi, src) = \sum_{i<j} s(i, j, \pi(i), \pi(j), src). \tag{7}$$

In their work, $s(i, j, \pi(i), \pi(j), src)$ is realized by a linear classifier, which is as shown in formula (8).

$$s(i, j, \pi(i), \pi(j), src) = f(i, j, \pi(i), \pi(j), src) * \theta. \tag{8}$$

Where, f is a feature vector while θ is corresponding feature weight vector. Under frame of the sequencing model, pre-sequencing problem on machine translation is transformed into a search process searching replacement with the highest score, which is as shown in formula (9).

$$\pi* = \arg\max_{\pi} s(\pi, src). \tag{9}$$

A key factor affecting the model performance is whether proper feature f can be designed. In work done by Tromble, he adopts a large number of high-dimension vocabulary characteristic. Because it is difficult to popularize vocabulary characteristic, they introduce coarse granularity characteristic, such as word category and pos tagging etc. for smoothness.

3. Dynamic Bayesian network reasoning and fuzzy comprehensive evaluation

3.1. Reason to select dynamic Bayesian network

Dynamic Bayesian network (DBN) is temporal extension of Bayesian network (BN). It possesses functional characteristics of static Bayesian network, and embodies effect of sample data on network structure in time domain in more accurate way. The method is applicable to evaluation to effect of situation element change on threat to the whole defensive system in network space combat information security defense situation evaluation. Integrate temporal causal relation between adjacent time slice into causal relation within the same time slice, and make dynamic analysis through quantitative reasoning[12], and dynamic Bayesian network can be simply defined as $(B_0, B_{\rightarrow})$, where B_0 is its BN in T_0 (time slice under original state), and prior probability $P(X_0)$ of hidden node and observation node can be obtained from BN structure, and $B_{\rightarrow}$ is figure constituted by BN in each time slice.

DBN possesses new knowledge integration, complete object expression, deduction and learning function, and has relatively good effect when making modeling analysis to uncertain problems having random process characteristic, and network structure of DBN is as shown in Fig.1.

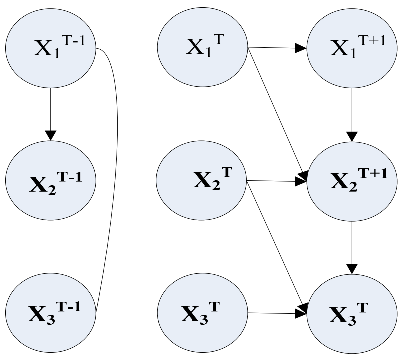

Fig. 1. Network structure of dynamic bayesian

3.2. Reasoning algorithm of dynamic Bayesian network

Reasoning algorithm of dynamic Bayesian network is derived from formula (1) Bayes formula:

$$p(x\,|y) = \frac{p(yx)}{p(y)} = \frac{p(yx)}{\sum_x p(yx)}\,. \tag{10}$$

The reasoning process is the same with that of static Bayesian network. Reasoning process of disperse static Bayesian network having n hidden nodes and m observation nodes can be reflected as mathematical process of formula (11) according to its conditional independence characteristic:

$$p(x_1, x_2, ..., x_n \,|y1, y2, ..., y_m\,) = \frac{\prod\limits_j p(y_j \,|p_a(Y_j)) \prod\limits_i p(x_i \,|p_a(X_i))}{\sum\limits_{x_1,x_2,...,x_n} \prod\limits_j p\,(yj\,|p_a(Yj)\,) \prod\limits_i p\,(x_i \,|p_a(X_i)\,)}. \quad (11)$$

In the formula, $i \in [1, n]$, $j \in [1, m[]$. x_i is 1 state value of X_i. $p_a(Y_j)$ is parent nodes set of Y_j.

When the quantity of hidden node and observable node is relatively few in network, or when coupling of node is relatively strong, and network structure layer is relatively few and comparatively few time slice is considered, each time slice of DBN can be regarded as 1 static Bayesian network, and in the case of gradual increase of node or node coupling increase, DBN constituted by T time slice can be obtained temporally, and its reasoning process can be reflected as formula (12):

$$\begin{aligned} & p(x_{11}, ..., x_{1n}, ..., x_{T1}, ..., x_{Tn} \,|Y_{11o}, Y_{12o}, ..., Y_{1mo}\,, ..., Y_{T1o}, Y_{T2o}, Y_{Tmo}) \\ & = \sum_{y_{11}y_{12}\cdots y_{Tm}} \frac{\prod\limits_{i,j} p(y_{ij} \,|p_a(Y_{ij})) \prod\limits_{i,k} p(x_{ik} \left| p_a(X_{ik})) \prod\limits_{i,j} p(Y_{ijo} = y_{ijo}) \right.}{\sum\limits_{x_{11},x_{21},...,x_{T1}\cdots x_{Tn}} \prod\limits_{i,j} p\,(y_{ij}\,|p_a(Y_{ij})\,) \prod\limits_{i,k} p\,(x_{ik}\,|p_a(X_{ik})\,)}. \end{aligned} \quad (12)$$

In the formula, $i \in [1, T]\,, j \in [1, m]\,, K \in [1, n]$. x_{ij} is 1 state value of X_{ij}; i is time slice; j is hidden node; y_{ij} is value of observational variable Y_{ij}; $p_a(Y_{ij})$ is parent nodes set of y_{ij}; Y_{ijo} is observational state of observable node j within the i^{th} time slice; $p(Y_{ijo} = y_{ijo})$ is membership of continuous observed values of Y_{ij} belonging to state y_{ij}.

3.3. Dynamic Bayesian network model for situation evaluation

2 kinds of node are used in situation evaluation to represent situation and event, and hierarchical relation between situation and event is embodied in model, which is as shown in Fig.2.

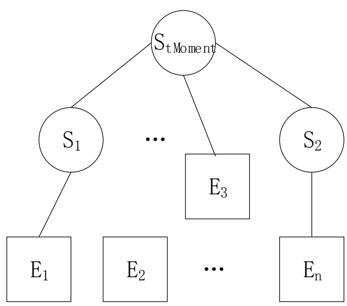

Fig. 2. Situation evaluation model of dynamic bayesian network

situation -situation,situation-eventevent-eventIn dynamic Bayesian network model established, round node represents single-element situation node, each node represents certain situation, and square node represents event node, and in BN, each node is connected through 3 kinds of relationship, i.e. situation –situation, situation-event, event-event.

①Situation–situation connection

If 1 situation node S is constituted by mutual independent situation subnode, $s_1, s_2, s_3 ... s_n$, their mutual relation can be expressed through single-layer tree structure:

$$Bel(S) = \sum_{i=1}^{n} Bel(s_i)\,. \tag{13}$$

In the formula, Bel is belief function, being sum of confidence coefficient of 1 situation node and situation confidence coefficient of mutual independent subnode, and confidence coefficient of child situation node is not greater than confidence coefficient of father situation node. If confidence coefficient of 1 situation is quite little, its child situation can be ignored, which can reduce the number of node contained in situation evaluation process.

② situation –event connection , Situation –event connection

The connection represents causal relation between situation and related event; if 1 event E_i contains n states, then probability matrix can be used, which is shown as follows:

$$P(E_i\,|S\,) = (P_{11,} ... P_{n1})^T\,. \tag{14}$$

③ Event-event connection

The connection represents logical reasoning relationship between different event nodes, and if event node E_w having a states and event E_k having b states are connected in BN, it can be expressed as follows through conditional probability matrix P_{ab} of $a \times b$:

$$P_{ab} = \begin{vmatrix} p_{11} & p_{12} & \cdots & p_{1b} \\ p_{21} & p_{22} & \cdots & p_{2b} \\ \cdots & \cdots & \cdots & \cdots \\ p_{b1} & p_{b2} & \cdots & p_{ab} \end{vmatrix}\,.$$

In the formula, $p_{ij} = p(E_{kj}\,|p_{wi}\,)$, which represents probability of state E_{kj} of event E_k if state of node E_w is E_{wi}.

3.4. Reason to select fuzzy comprehensive evaluation method

Fuzzy comprehensive evaluation analyzes complex fuzzy system through fuzzy transformation principle, and the method is frequently applied to multiple attribute decision making problems. Comprehensive decision is made to problem by making qualitative and quantitive analysis and fuzzy evaluation to a great number of complex influential factors. In medical English-Chinese interpretation, situation evaluation index set can be considered as 1 multi-index evaluation problem, and multilevel and multi-factor comprehensive analysis shall be made to index element established, and situation index of evaluation network constructed at each level is qualitative description with high complexity, so it is applicable to the method.

4. Experimental analysis

4.1. Experimental data

Our experimental data is divided into three parts: the first part is monolingual text for vocabulary vector representation and language model training; the second part is bilingual data for pre-sequencing model and translation model training; the third part is test data for translation effect evaluation.

(1) Monolingual corpus: our monolingual text is monolingual text collected from Internet. After normalization and deduplication treatment etc., we obtain English text of 1 billion sentences, Chinese text of 0.4 billion sentences, and Japanese text of 0.2 billion sentences roughly, of which English is used for language model training as target language; Chinese and Japanese are used for vocabulary vectorization representation training as source language.

(2) Bilingual parallel corpus: our parallel corpus is collected from Internet automatically. In the experiment, data used by us from Chinese to English includes 26 million sentence pairs and data from Japanese to English includes 15 million sentence pairs roughly. We use these pre-sequencing models for corpus training and translation models.

(3) Machine translation experiment test data: we adopt standard NIST machine translation evaluation test set for Chinese to English, of which NIST05 is taken as development set while NIST06 and NIST08 are taken as test set. We adopt news corpus of 5000 sentences translated by human for experiment from Japanese to English, of which 2500 sentences are taken as development set while 2500 sentences are taken as test set.

4.2. Measurement to pre-sequencing result

In addition to machine translation result, we want to measure performance of pre-sequencing in word sequencing task. Therefore, we adopt word alignment cross connection number between source language and target language for evaluation. The more consistent the word order of source language and target language after sequencing is, the smaller the word alignment cross connection number between them shall be, and the better the pre-sequencing effect will be. Because there may be mistake in word alignment generated automatically, we choose 500 sentence pairs from Chinese-English and Japanese-English dataset for word alignment label, and make test on the 500 sentence pairs.

Seen from Table 1, pre-sequencing helps to decrease word alignment cross connection number. In Japanese-English dataset, pre-sequencing has obvious improvement effect on word order. In Chinese-English dataset, pre-sequencing also has certain effect, and pre-sequencing model based on dynamic Bayesian network has better effect than pre-sequencing model based on sparse feature. These experimental results are consistent with experimental result tendency of machine translation performance: system with relatively small word alignment cross connection number has relatively better translation performance.

Table 1. Average value of alignment cross connection number of each sentence pair

Experimental system	Word from Chinese to English	Word from Japanese to English
No PR	30.3	70.9
Sparsc PR	25.6	35.7
NN PR	17.0	34.0

4.3. Comparison with other pre-sequencing method

In addition to be considered as linear sequencing (LO) problem, pre-sequencing can also be formalized as asymmetric traveling salesman (ATS) problem or general ranking problem. We realize pre-sequencing system of asymmetric traveling salesman method based on sparse feature and general ranking method, and experimental results are as shown in Table 2∼3.

Table 2. Comparison on different pre-sequencing methods from Chinese to English

Experimental system	NIST05 (development)	NIST06	NISTO8
ATS PR	0.417	0.378	0.313
Ranking PR	0.415	0.375	0.311
LO PR	0.416	0.376	0.314
LO-NN PR	0.422	0.382	0.320

Table 3. Comparison on different pre-sequencing methods from Japanese to English

Experimental system	Development set	Test set
ATS PR	0.243	0.242
Ranking PR	0.242	0.239
LO PR	0.245	0.241
LO-NN PR	0.246	0.242

From experimental result, we find that results of different formalized pre-sequencing models are quite close. In experiment from Japanese to English, results of 3 kinds of pre-sequencing model based on sparse feature and pre-sequencing model based on dynamic Bayesian network are close; in experiment from Chinese to English, results of 3 kinds of pre-sequencing model based on sparse feature are close, but results of linear sequencing model based on dynamic Bayesian network are superior to results of model based on sparse feature only, which shows that in our experiment, effect difference of different formalization of pre-sequencing model is not large under the same input characteristics.

5. Conclusion

This paper proposes a kind of pre-sequencing model based on dynamic Bayesian network for statistical machine translation. Utilizing Bayesian language model method, this paper learns vocabulary vector representation from unlabeled text, and combines the vocabulary representation with other features and integrates into a linear sequencing model by utilizing a multilayer dynamic Bayesian network. Results of experiment from Chinese to English and from Japanese to English show that compared with baseline system, pre-sequencing model that is proposed in this paper and that is based on dynamic Bayesian network can improve performance of machine translation system obviously.

Along current direction, we intend to explore vector representation method of phrase and effect of the representation on pre-sequencing model of machine translation in the future. In addition, we plan to study expression of method in this paper in system based on syntax and research how to make vector representation learning to more abstract syntax tree fragment.

References

[1] J. J. Faig, A. Moretti, L. B. Joseph, Y. Y. Zhang, M. J. Nova, K. Smith, and K. E. Uhrich: (2017). *Biodegradable Kojic Acid-Based Polymers: Controlled Delivery of Bioactives for Melanogenesis Inhibition*, Biomacromolecules, 18(2), 363-373.

[2] Z. Lv, A. Halawani, S. Feng S., H. Li H., & 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.

[3] Y. Z. Chen, F. J. Tang, Y. Bao, Y. Tang, G. D. Chen: *A Fe-C coated long period fiber grating sensor for corrosion induced mass loss measurement.* Optics letters, *41* (2016), 2306–2309.

[4] Y. Du, Y. Z. Chen, Y. Y. Zhuang, C. Zhu, F. J. Tang, J. Huang: *Probing Nanostrain via a Mechanically Designed Optical Fiber Interferometer.* IEEE Photonics Technology Letters, *29* (2017), 1348–1351.

[5] 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.

[6] Y. Y. Zhang, Q. Li, W. J. Welsh, P. V. Moghe, and K. E. Uhrich: *Micellar and Structural Stability of Nanoscale Amphiphilic Polymers: Implications for Anti-atherosclerotic Bioactivity*, Biomaterials, *84* (2016), 230–240.

[7] J. W. Chan, Y. Y. Zhang, and K. E. Uhrich: *Amphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation*, Bioconjugate Chemistry, *26* (2015), No. 7, 1359–1369.

[8] Y. J. Liao, F. L. School, H. N. University: *Fuzzy Language and its Aesthetic Sense in Literary Translation: A Case Study of The Border Town*[J]. Overseas English (2013).

[9] D. Muhammad: *Characterisation of Arsenic Distribution in the Contaminated Sediments Using Principal Component Analysis based on The Four-Step Extraction Protocol*[J]. Foreign Language Education (2008).

[10] L. Zhou, L. U. Xiu-Ying: *Pragmatic Analysis of Fuzzy Language in Newspaper English and Its Translation Strategies*[J]. Journal of East China Jiaotong University (2008).

[11] X. N. He: *On Fuzziness in English and Chinese and its Corresponding Translation Strategies*[J]. Journal of Tonghua Normal University (2009).

[12] A. Cumbers, D. Mackinnon, R. Mcmaster: *Institutions, Power and Space Assessing the Limits to Institutionalism in Economic Geography*[J]. European Urban & Regional Studies, *10* (2003), No. 4, 325–342.

Received May 7, 2017