# A method of mining the change of word semantic across different timescales

**Maoyuan Zhang[1, a], Shuyuan Sun[2, b] and Yibo Wang[3, c]**

[1,2,3]School of Computer, Central China Normal University, Wuhan, 430079, China

[a]yibowong@qq.com, [b]874799867@qq.com, [c] panxiaohang_love@126.com

**Keywords:** Word variation, Word embeddings, Cloud model.

**Abstract.** Language evolves over time, and language changes its meaning due to cultural changes, technological inventions, or political events. Language variation and change is an important branch of sociolinguistics and has achieved remarkable achievements. But there is seldom study conducted from the aspect of the natural language processing. In this paper, we propose a probabilistic language model of timestamp text data, which can track the semantic evolution of a single word over time and use the cloud model to compute the change of words. Experiments show that the proposed method has good semantic information and useful results in semantic change and change analysis.

## 1. Introduction

Language is developed continuously, as a component of language, the words have the development and change of the language at the same time. So, the meaning of words is relatively stable, but it is also gradually developing. Language evolves over time and words change their meaning due to cultural shifts, technological inventions, or political events. The language variation and change has been an important subfield in sociolinguistics, and has made remarkable achievements [1]. However, linguists generally choose empirical investigation (which may be a qualitative or quantitative analysis), this is usually laborious and time-consuming, rarely from Natural Language Processing (NLP) perspective. In NLP, the methodology is a typical corpus based statistical method which relies on the context (lexical or syntactic) of the target words and gives their statistical trends in semantic or usage [2].

In recent years, distributed word representations or embeddings has been proved a powerful tool for modeling semantic relations between individual words. Word embeddings simulate the distribution of words based on their surrounding words in a training corpus. Word embeddings are currently formulated as a static model, which assumes that any given word has the same meaning in the entire text corpus. Bengio et al.[3] used the $t-n+1$ words to the $t-1$word as the regression neural network input, the T words as output regression neural network. Mikolov et al. [4]proposed the skip-gram model with negative sampling (word2vec) as a scalable word embedding approach that relies on stochastic gradient descent.

In this paper, we proposed a probabilistic state space model which allows us to share information across all times. The distributed word representations are used to map the word semantic to the multidimensional space, and the similarity computation is used to find the similar word in the space. Then, we convert the word and its related words into three characteristic values of the cloud model. For the same word in different time periods, the degree of variation can be measured by calculating the overlap degree of cloud model.

Our paper is structured as follows. In section 2 we presents our method specifically and show experimental results in section 3. The last section concludes this paper and discusses the future work.

Copyright © (2017) Francis Academic Press, UK

## 2. Word Semantic Variation Mining based on Method

### 2.1 Skip-Gram Model

The Skip-gram model proposed by Mikolov et al. can train word embedding quickly and efficiently from large, unlabeled text data set. The skip-gram model uses the sliding window to capture the co-occurrence information of words and generates high-dimensional distributed word vectors for each word, so that the generated word vectors have semantic and grammatical information and its structure is shown in Figure 1.

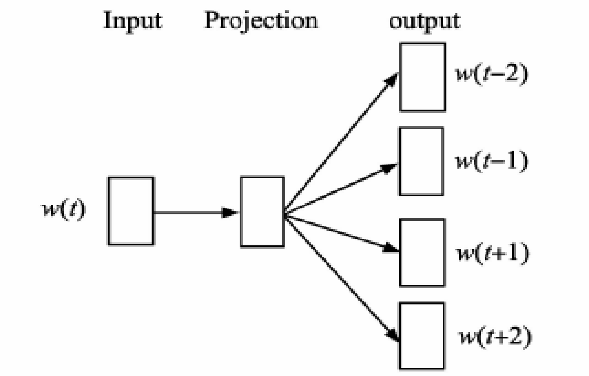

Fig.1. The skip-gram model

As shown in Figure 1, the skip-gram model consists of three layers: the input layer, the projection layer and the output layer. The objective function $L$ is described as Eq.(1) and Eq.(2):

$$L = \sum_{w \in V} \log p(W_c \mid w_i) \tag{1}$$

$$p(W_c \mid w) = \prod_{w_j \in W_c} p(w_j \mid w_i) \tag{2}$$

The skip-gram model is a probabilistic version of word2vec.For each word $i, j$ in the vocabulary, the model assigns probabilities that word $i$ appears in the context of word $j$. The generative model assumes that many word-word pairs $(i, j)$ are uniformly drawn from the vocabulary and tested for being a word-context pair; hence a separate random indicator $z_{ij}$ is associated with each drawn pair.

For the word vector $v_i, v_j$ of $w_i$ and $w_j$ ,we use the euclidean distance to measure the sematic similarity , and the calculation is described as Eq.(3)

$$Sim(w_i, w_j) = \left\| v_i - v_j \right\|_2 \tag{3}$$

Where $Sim(w_i, w_j)$ is the sematic similarity between word $w_i$ and word $w_j$ , $v_i, v_j$ is the word vector of word $w_i$ and word $w_j$.

### 2.2 Cloud Model

Unlike cloud computing, cloud model[5] is the cognitive model of Natural Language Processing. After a series of cognitive science studies, artificial intelligence and knowledge representation, Deyi Li proposed the cloud model theory. From a natural language point of view, it is more in line with the nature of things than traditional mathematics. The basic theories such as fuzzy theory, probability theory and rough set theory provide the basis for the establishment of cloud model. It describes the uncertainty of linguistic concepts, especially randomness and fuzziness, and achieves the uncertainty transformation between linguistic concepts and quantized values. A word and its relation form a concept. In order to express concepts, the three quantized cloud model is described below.

(1)Concept Expectation: Concept Expectation is the mathematical expectation of the relative words belonging to a concept in the universal. It can be regarded as the most representative and typical sample of the qualitative concept.

(2) Concept Entropy: Concept Entropy represents the uncertainty measurement of a qualitative concept. It is determined by both the randomness and the fuzziness of the concept. As the measurement of randomness, it reflects the dispersing extent of the relative words. On the other hand, it is also the measurement of fuzziness, representing the scope of the universe that can be accepted by the concept.

(3) Hyper entropy: Hyper entropy is the uncertain degree of entropy.

Their equations are introduced as Eq.(4):

$$\begin{cases} Ex = \frac{1}{l}\sum_{i=1}^{l} Xi \\ En = \frac{1}{l}\sqrt{\frac{\pi}{2}}\sum_{i=1}^{l}\left|Xi - Ex\right| \\ He = \sqrt{\left|S^{2} - En^{2}\right|} \end{cases} \tag{4}$$

where Ex is the value of concept expectation, En is the value of concept entropy, He is the value of Hyper entropy, $S^{2} = \frac{1}{l-1}\sum_{i=1}^{l}\left(Xi - Ex\right)^{2}$ ,l is the total number of relative words of center word, Xi is the relevance between the center word and its ith relative word . A concept can be expressed as C(Ex;En;He).

Considering the internal relations of Ex,En and He,the region Rs[Ex-En;Ex+En], Rs[Ex-2En;Ex+2En],Rs[Ex-3En;Ex+3En]contributes 68.28%,95.46%, 99.74% to the meaning of concept.When two words express or indicate approximately similar meaning, their quantification will be much near each other, so the rate of intersection over the whole two numerical regions reflects the similarity of two words. The higher the rate, the more similar the word pair is.Suppose vector C1(Ex1,En1,He1) stands for word w1, vector C2(Ex2,En2,He2) represents sentence w2, the region Rc [Ex -3En ;Ex+3En ] represents concept C.Finally the score between C1 and c2 derived from Eq.(5) represents the variation degree of the word pair.

$$R_{SC}(c1,c2) = 1 - \frac{R_{C1}\bigcap R_{C2}}{R_{C1}\bigcup R_{C2}} \tag{5}$$

## 3. Experiment and Results

In this paper, we use data collected from network news, and we segment all text automatic segmentation software using ICTCLAS to do all data preprocessing, such as stop words. Then we use skip-gram to model the text. Therefore, these words will be represented as continuous vectors, in which the weights of vectors are computed directly in order to maximize the probability of the context in which the simulated words appear. This allows an efficient representation of the model trained on a relatively small number of large amounts of data. We use the 200 dimensional vector on the 100 million word Sina News dataset, which covers more than 180 words and phrases, which is a fairly large vocabulary. For each entry, we use cosine similarity to compute twenty of the most similar entries. We calculated the extent of all the changes in the word and took the first ten. Table 1 shows the top 10 sensitive words.

Table 1 The top 10 sensitive words

| Ranking | Word | semantic change | Ranking | Word | semantic change |
|---|---|---|---|---|---|
| 1 | passwords | 0.89123 | 6 | consume | 0.85388 |
| 2 | giants | 0.87621 | 7 | arguing | 0.89567 |
| 3 | donation | 0.96382 | 8 | yankees | 0.84726 |
| 4 | partitions | 0.70541 | 9 | intelligence | 0.88731 |
| 5 | greeting | 0.83281 | 10 | party | 0.88562 |

## 4. Conclusions

This paper studies the semantic change and change mining from the calculation of lexical semantics, preliminary experiments show that our method achieves the semantic change trend and word level features of word analysis help results. Our future work will focus on the following aspects: firstly, the use of finer algorithmic process subjects in semantic mining of words leads to finer model design. Secondly, many other historical changes mining can be conducted based on the corpus.

## 5. Acknowledgments

This work was supported by Humanity and Social Science Youth Foundation of Ministry of Education of China (No. 15YJC870029), Research Planning Project of National Language Committee (No.YB135-40), the self-determined research funds of CCNU from the colleges' basic research and operation of MOE(No. CCNU16A02049, No. CCNU16A06039).

## References

[1] Walker J A. The Handbook of Language Variation and Change (review)[J]. Language, 2004, 80(3):591-594.

[2] Jurafsky D, Martin J H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition[J]. Prentice Hall, 2008, 26(4):638-641.

[3] Bengio Y, Schwenk H, Senecal J S,et al. Neural Probabilistic Language Models[J]. Journal of Machine Learning Research, 2006, 3(6): 1137-1155.

[4] Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado,Greg S, and Dean, Jeff. Distributed Representations of Words and Phrases and their Compositionality. In Advances in Neural Information Processing Systems 26, pp.
3111–3119. 2013.

[5] Li D, Liu C, Gan W. A new cognitive model: Cloud model[J]. International Journal of Intelligent Systems, 2009, 24(3):357-375.