IALP 2019, Shanghai, Nov 15-17, 2019

Diachronic Synonymy and Polysemy: Exploring Dynamic Relation Between Forms and Meanings of Words Based on Word Embeddings

Shichen Liang § †, Jianyu Zheng‡, Xuemei Tang § †, Renfen Hu § †, Zhiying Liu § †
§ Institute of Chinese Information Processing, Beijing Normal University, Beijing China
†UltraPower-BNU Joint Laboratory for Artificial Intelligence, Beijing China
{shichen, tangxuemei, irishu}@mail.bnu.edu.cn, liuzhy@bnu.edu.cn
‡Advanced Innovation Center for Future Education, Beijing Normal University, Beijing China
zheng_jianyu@126.com

***Abstract*—In recent years, there has been a large number of publications that use distributed methods to track temporal changes in lexical semantics. However, most current researches only state the simple fact that the meaning of words has changed, lacking more detailed and in-depth analysis. We combine linguistic theory and word embedding model to study Chinese diachronic semantics. Specifically, two methods of word analogy and word similarity are associated with diachronic synonymy and diachronic polysemy respectively, and the aligned diachronic word embeddings are used to detect the changes of relationship between forms and meanings of words. Through experiments and case studies, our method achieves the ideal result. We also find that the evolution of Chinese vocabulary is closely related to social development, and there is a certain correlation between the polysemy and synonymy of the word meaning.**

Keywords—Diachronic Synonymy, Diachronic Polysemy, Lexical Semantic, Word Embedding

I. Introduction

In various subsystems of the language, vocabulary system holds the fastest changes. As the most basic unit of vocabulary system, a word is a relative unity of form and meaning. With the development of society and cultural communication, the connotation and denotation of a word's meaning is changing. On the one hand, the same word will hold different meanings That is, old words are with new meanings; On the other hand, the same concept will be also expressed by different words. Faced with that situation, we combined word embeddings and some theories in linguistics, and explored the evolution of lexical semantics from a diachronic perspective.

In recent years, with the boom of deep learning, the word embeddings which can vectorize vocabularies are also used extensively. As each word is represented distributely, the semantic relationship between words can be examined by distances between vectors. For example, as to the following two pairs of words: "Dad" - "Mom", "Dad" - "Apple", by calculation, the similarity of the previous word pair is much higher than the later, which is consistent with people's cognition. In addition, word embeddings can also be used for semantic analogy between words. That is, given the first three words, the fourth word can be automatically inferred by calculations among their vectors. The most famous example is: king + man – woman = queen. Because word embeddings can describe the words' meanings to some extent, they are used in many tasks of natural language processing.

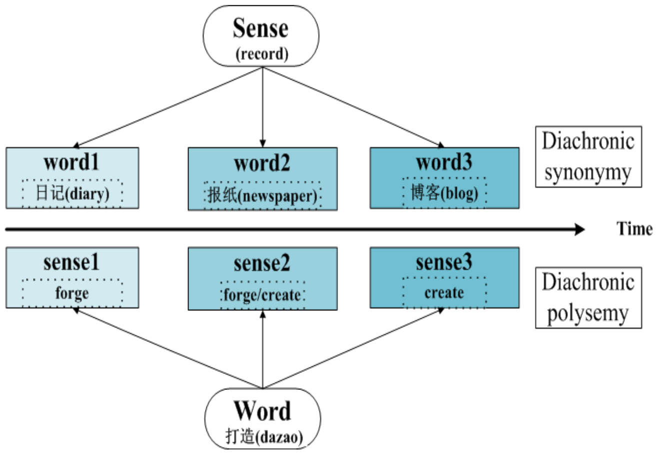

Fig. 1. A sketch map about dynamic relation between forms and meanings of words. As shown in the upper half of the figure, the word “blog” does not exist in the past. The meaning contained in the word “blog” was represented by “diary” in the past period, so the two words hold the semantic equivalence from the diachronic perspective.

In this paper, we study the relation between forms and meanings of words from a diachronic perspective based on the word embeddings. Firstly, according to the diachronic corpus, word embeddings during different periods are respectively trained, and the vector spaces are aligned by a certain alignment method. Then, from the perspectives of diachronic synonymy and diachronic polysemy, this paper explores the dynamic relation between forms and meanings of words. As shown in figure1: from the perspective of diachronic synonymy, it is possible to observe whether the same meaning is represented by different word forms in different periods by the diachronic word analogy; And from the perspective of diachronic polysemy, whether the same form has different meanings in different periods by the diachronic word similarity. The main contributions of this paper are as follows:①the word analogy and word similarity are applied to the study of diachronic semantics, which is different from the previous work, using these two methods to evaluate the quality of word embeddings;②combining the two perspectives of diachronic synonymy and diachronic polysemy, we study the dynamic relation between forms and meanings of words.

II. Related Works

The evolution of word meaning is a common phenomenon in all languages, and has always been one of the central topics concerned by linguists[1]. Initially, researchers inferred the changes of word meaning by subjective language usage preferences. Later, with the establishment of large-scale corpus, scholars carried out the research by the "data-driven" approach from a quantitative perspective[2]. Specifically, by using the corpus to query how a word is used in the context, the evolution of a word's meaning can be summarized[3][4][5][6]. Among those, counting the usage frequency of a certain word under the diachronic condition is the most typical method[7] [8][9]. However, it also has many shortcomings, such as the fialure in characterizing and measurement of the relative semantic relationship with other words.

With the deep learning being widely used in natural language processing, word embedding has become an effective way to represent words quantitatively. The word embedding model was originally proposed by Mikolov[10]. This model can learn unsupervised low-dimensional dense representations of each word from large-scale texts. Later, researchers applied this technique to explore the evolution of word meaning[11][12][13][14][15] . Kim et al. [16]investigated the semantic variation of certain words by calculating the cosine similarity between word embeddings, such as the words "cell" and "gay". It is worth to mention that they used incremental updates and Continuous Skip-gram with negative sampling (SGNS) to improve the representation of the word. By calculating the similarity between word embeddings, Kulkarn et al.[17]observed the neighbor words of a given word in different periods, and judged whether the meaning of word has evolved. Later, from the perspective of word analogy, Szymanski[18]first proposed Temporal Word Analogies to induce the corresponding analogical relation between words during different periods, such as "Ronald Reagan in 1987 is like Bill Clinton in 1997".

In summary, the study of diachronic meaning by word similarity has gained some achievements. But there is relatively little work in word analogy, especially in Chinese. At the same time, there is no study to analyze lexical semantics macroscopically by combining forms and meanings of words. Inspired by structural linguistics[19], this paper takes roots in the evolution of lexical semantic systems from two perspectives: diachronic synonymy and diachronic polysemy. At the same time, we apply the word analogy and word similarity of word embedding to obeserve the two linguistic phenomena, and analyze the dynamic changes of lexical semantics.

III. TRAINING DIACHRONIC WORD EMBEDDINGS

Before carrying out our experiment and analysis, we need to train diachronic word embeddings based on the corpus from different periods, so that we can observe the evolution relation between word meaning and word form in a diachronic level. Essentially, diachronic word embeddings are a set of aligned synchronic word embeddings. In this section, we constructed word embeddings in each decade from 1947 to 2017 with word2vec, and then aligned those vector spaces to get comparable diachronic word embeddings.

A. Word2vec

Word embeddings are low-dimension word representations trained from unlabeled corpora through neural models. According to Firth (1957)[20], distributional hypothesis is a theory claims that words appearing in similar contexts must have similar meanings and representations. This hypothesis is one of the theoretical bases of word embedding model.

At present, the word representation methods most commonly used include PPMI, SVD, Word2Vec, Glove and so on. The embeddings trained by Word2vec are optimized to predict co-occurrence relationships. Especially, given the input word, skip-gram model can predict the context information which is more suitable for studying semantic change. For this reason, we finally adopted the skip-gram model to construct word embeddings...

B. Datasets, pre-processing, and hyper-parameters

We collected the news in People's Daily from 1948 to 2017 to construct a corpus, as this newspaper is a national print media with the largest scale and the highest authority in China. All news spans 70 years, whose total size is 4.43GB. Taking social realities and experimental operability into account, we divided the data by 10 years, so there were 7 different vector space models.

Years	Tokens	Size(KB)
1948-1957	116382	430,786
1958-1967	157822	274,846
1968-1977	138431	451,674
1978-1987	150195	449,358
1988-1997	175085	354,597

TABLE I. STATISTICS OF 7 VECTOR SPACE MODELS.

Those embeddings in each period were trained by skip-gram with vector size 300, window side 5, min-count 5. As Table I shows, each model contains more than 100k tokens.

C. Alignment of vector space models

During training, those vectors are usually randomly initialized. Although the mutual relations between vectors may be consistent, due to the stochastic processing, vectors not trained in the same run are not comparable with each other. For this reason, It requires us to align each independent vector space model. To address this problem, previous studies have provided three approaches: Non-random initialization, Local linear regression and Orthogonal Procrustes.

In this work, we followed Terrence Saymanski (2017)[18]'s method: we assumed there was always a linear transformation can align two vector space models, and the meanings of most words did not change. First, we found the common words of two vector spaces and sampled them in proportion; Then, the results of sampling were respectively used as independent variables and dependent variables for linear regression, and the linear regression model was obtained by minimizing the distance (mean square error) of the two sampling spaces; Last, we applied the linear regression model to the entire vector space models.

A small change was adapted during random sampling. What we did wass not randomizing completely. According to Hamilton (2016)[21], words with higher word frequency have less change in meaning. Therefore, we made use of the HSK (an examination for testing the Chinese language level of foreigners) vocabularies issued by HanBan of China, which contains 8650 common words in Chinese. In order to get a more reasonable initial sampling, we reduced the proportion of random samples, and added 8650 words mentioned above to the set of random words. Through this way, we constructed Chinese diachronic word embeddings.

IV. RESULTS AND ANALYSIS

We explored the evolution relation between form and meaning of Chinese vocabularies from two perspectives: diachronic polysemy and diachronic synonymy. The specific measures are diachronic word analogy and diachronic word similarity. Previously, word similarity and analogy were methods for evaluating the synchronic word embeddings. Meanwhile, they are also useful tools to reveal relation of lexical semantic. In this section, we capture word sets which convey the same meaning in different periods by diachronic word analogy. At the same time, we detect words which hold different meanings in different periods by diachronic word similarity.

A. Analogy and diachronic synonymy

With replacement of things and concepts, the same concept may be expressed in different word forms in the diachronic level. Diachronic word analogy is a technical measure to establish a synonymous relationship between those words which have unrelatedness with each other at the synchronic level. In general, analogy is to infer the unfilled word based on the relationship between two word pairs, i.e. word W1 to word W2 is equivalent to word W3 and word W4. When it comes to diachronic word analogy, that is word wα at time tα is equivalent to word wβ at time tβ, which has a slight difference with traditional word analogy. For example, “微博”（Weibo, microblog）in years of 2008-2017 is equivalent to “墙报”（qiangbao, wall newspaper）in years of 1948-1957.

In this section, we select the vector space of a specific period as the basic vector space, then align the vector spaces of the other six periods with it through the alignment method mentioned in section III.C. When given a word in the basic vector space, we carry out word analogy to find diachronic synonyms in other periods. What we need to do is to find vectors closest to the word in other six vector spaces, and get the words represented by the corresponding vectors as a result of analogy. At last, there are seven words in each set of analogy. Ideally, if the correspondence between words and meanings does not change, there should be seven identical words in each set. Therefore, if different words are used for the same meaning at different times, then the diachronic word analogy will be able to discover a semantic consistency between these words. It is worth noting that the meaning here is a relatively broad definition. For example, Xi and MAO are different names which do not point to the same person, but they share a common concept, namely Chinese President. In cases like this, we also view them as two words that convey same meaning.

TABLE II. EXAMPLES OF WORDS FROM 2008-2017 AND THEIR ANALOGOUS OVER TIME*

1948-1957	1958-1967	1968-1977	1978-1987	1988-1997	1998-2007	2008-2017
王府井大街 WangFuJing Street	餐馆 restaurant	书摊 bookstall	赌场 casino	柜台 counter	专卖店 specialty store	淘宝 Taobao
书信 Letter	信件 letter	传真 fax	信函 letter	传真 fax	电子邮件 e-mail	电子邮件 e-mail
黑板 blackboard	卖报 newspaper	名片 business card	电话 phone	手机短信 SMS	电子邮件 e-mail	微信 WeChat
墙报 wall newspaper	副刊 supplement	栏目 column	新闻媒体 media	人民网 People's Network	网站 website	微博 microblog
日记 diary	日记 diary	日记 diary	书评 book review	黑板报 blackboard	报纸 newspaper	博客 blog
顾客 customer	代售 sale	烟酒 tobacco& alcohol	小卖部 commissary	送货上门 delivery	售货员 salesperson	外卖 take-out
航线 air line	电气化铁路 railway	高速公路 highway	高速公路 highway	高速公路 highway	高速公路 highway	·高铁 high-speed railway

*The last column (2008-2017) is the words in the reference period, and the vector spaces of the other six periods are respectively aligned with this period to find the words that is closest to words of 2008-2017. This table shows the result of word analogy, for example, “博客”(blog)+timeα（2008-2017）= “日记”（diary）+timeβ（1948-1957）

We selected 50 neologism which are commonly used in 2008-2017 to constitute a small vocabulary for analogy. Those words include political words, life words, ideology, famous figures, institution, and we eventually form 300 pairs of word analogies. These new words from 2008 to 2017 are respectively compared in six other periods to get the synonyms in those periods. The results show that political words and life words have achieved a better performance. As table II shows, the analogy of life vocabularies roughly shows the development of Chinese society, indicating that the diachronic word analogy is effective to some extent. For example, the term "blog" appeared in the Internet era, also known as "network diary". As the result shows, in 1948-1977, before the Internet became popular, the corresponding word of this concept is exactly "diary".

B. Similarity and Diachronic polysemy

In this section, we observe old words with changeable meanings and study how those changes happen by diachronic word similarity. This issue can be discussed from two aspects. When a word holds different meanings in the diachronic level. On the one hand, the word's similarity with itself in previous period will decrease; and on the other hand, its similarity with other words will also change. The details are as follows：

1) Diachronic similarity of individual words.

The similarity between a word and itself in previous period is an important reference standard to measure the degree of meaning change. If the lower the diachronic similarity of a word is, the more remarkable its diachronic polysemy.

By the vector space model after alignment, we calculate the cosine similarity of the vector (V_i, V_{i+1}) of each word (W) in every two periods (t_i, t_{i+1}), then sum and average the results of similarities. In this way, we get diachronic similarities of words from 1948 to 2017:

$$\text{similarity} = \frac{1}{n}\sum_{t=1}^{n-1}\frac{V_t \cdot V_{t+1}}{||V_t|| \cdot ||V_{t+1}||} \qquad (1)$$

n indicates the number of periods, here n=7.

TABLE III. SOME EXAMPLES OF DIACHRONIC WORD SIMILARITY.

words	diachronic word similarity
打造(forge; create)	0.132075
曝光(expose; make sth. public)	0.205325
登录(land; log in)	0.325444
家教(domestic or family education ; private teacher)	0.363201
绿色(green; environmentally friendly)	0.460422
变脸(face changing; turn hostile suddenly)	0.482308
赞(praise; favor)	0.493929
强暴(violent; rape)	0.549233
充电(charge; study)	0.549825
漫游(travel around; roam)	0.596138
阳光(sunlight ; cheerful ; transparent)	0.647266

In addition, we have collected some words from previous linguistic studies that did show diachronic polysemy. Unlike traditional linguistic research methods, word similarity can quantify the degree of change in meaning. Some examples are shown in Table III. All of these words are diachronic polysemes, but the competitiveness of the new meaning leads to the content of similarity. For example, the word "打造" (forge; create)with relatively low similarity originally refers to "forge", it now becomes similar to "create". Two meanings of this word are so different and the original meaning was almost lost; the similarity is slightly higher for the words "家教" (domestic or family education ; private teacher) and "透明"(transparent ; non-secretive). Those words have two meanings in parallel, and the two meanings are both used at the same period, so the degree of change is moderate; The word "阳光"(sunlight ; cheerful ; transparent) has three different meanings, and the most primitive meaning (sun's rays)is still the most frequently used. So it has relatively high similarity.

2) Diachronic similarity of multiple words

The phenomenon of diachronic polysemy can not only be reflected from the diachronic similarity of words but also from the change of synchronic synonymy of words. When the a word's meaning changes, it means that its synonyms of the word in previous period are broken, while the synonyms in the new period are established. Therefore, we can examine the diachronic polysemy and the movement direction of a word by examining the similarity of words to their neighbors at different times.

According to previous research and Modern Chinese Dictionary, we obtained 20 pairs of words: as the time passing, the meaning of one or two in those word pairs changes, and the relative distance between the pair is also altered. In other words, the semantic relationship between them could be changed. Semantic shifts can be observed as the target word moves closer to or further apart from its neighbors. We measure the movement of words by calculating the cosine similarity between word pairs over time.

If the similarity between W_1 and W_2 increases from time i to time j(j>i), it indicates that the meanings of W_1 and W_2 are getting closer and closer. Vice versa, the gap between the two meanings is getting narrower. We assume that the meaning of W_2 has not changed, then the meaning of W_1 must have changed. That is, W_1 is a diachronic polysemy. For example, according to the first column in Table IV, there are words whose meaning have changed, and the words in the second and third columns hold stable meanings. Through the changes of similarity of multiple sets of word pairs, we can determine the direction of meaning change. For example, the word "破产"(bankrupt; fail）listed in the Table IV originally refers to the property problem of an individual or a company, after then it captures a metaphorical usage means a failure.

Finally, we find that the polysemy and synonymy of words actually change synchronously. If a word in word pairs changes its meaning, its temporal similarity to other words will also be changed. Conversely, if the diachronic similarity between the word pair changes greatly, it is likely that the meaning of one word in the pair has changed. The change of the meaning is reflected in the relationship between words. The existence of diachronic polysemy often reflects the change of word's synonymic relation in the diachronic level. That is to say, from the diachronic perspective, the polysemous relation and the synonymy relation of words are interrelated.

TABLE IV. SEMANTIC SHIFTS CAPTURED BY HISTORICAL WORD EMBEDDINGS

Words	Moving closer	Moving away	Shift time
钓钩(hook)	骗局(fraud)	钓鱼(fishing)	1998-2007

阶级(steps; class)	阶层(stratum)	台阶(steps)	1948-1957
破产(bankrupt; fail)	失败(fail)	欠债(owe a debt)	1988-1997
包袱(bundle; joke)	笑料(joke)		1978-1987
程序(procedural; programme)	电脑(computer)		1988-1997
纠结(intertwine; confused)	烦闷(depressed)		1978-1987
萌(sprout; lovely)	可爱(lovely)		1998-2007
粉(powder; fans)	粉丝(fans)		1998-2007
霹雳(thunderbolt;unexcepted event)	坏消息(bad news)		1978-1987
晒(bask; share)	秀(show off)		1998-2007
山寨(village; copycatting)	假(fake)		2008-2017
小姐(miss; prostitute)	妓女(hooker)		1998-2007
宅(house; stay at home)	呆(stay)		2008-2017
总裁(director-general; chairman)	总经理(GM)		1978-1987

V. CONCLUSION

Based on the diachronic word embedding model, this paper uses word similarity and word analogy to investigate the dynamic relation between word form and semantics over time. According to the experiment results from some typical cases, our method has achieved ideal performance. In diachronic word analogy, we can find the corresponding words of the new words in the past, which can realize the semantic continuity between new and old vocabularies; In diachronic word similarity, we can detect the generation and development of the new meaning of the old words, which can achieve the semantic continuity of the same words. In addition, we also find that the lexical semantic change of Chinese vocabularies is closely related to the development of Chinese society.

However, as to the same word form, there is only a vector representation trained by word2vec generally, so the multiple meanings of the words could not be distinguished, and we could not explore the semantic changes thoroughly. In addition, the BERT model risen recently can also better deal with the problem of polysemous words, and we expect the idea in this paper can achieve more performance on BERT.

REFERENCES

[1] Baugh, A. C.& T. Cable. A History of the English Language. London: Routledge，1993.

[2] Elizabeth Traugott. Semantic change. Oxford Research Encyclopedias: Linguistics,2017.

[3] M. Hilpert. Germanic future constructions: A usage-based approach to language change. Benjamins, Amsterdam, Netherlands,2008.

[4] Stefan Th. Gries. 1999. Particle movement: a cognitive and functional approach. Cognitive Linguistics, 2008, pp. 105–145.

[5] D. Kerremans, S. Stegmayr, and H.-J. Schmid. The NeoCrawler: Identifying and retrieving neologisms from the internet and monitoring ongoing change. In K. Allan and J. A. Robinson,editors, Current methods in historical semantics, De Gruyter Mouton, 2010, pp. 130–160.

[6] Kristina Gulordava and Marco Baroni. A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics, Edinburgh, UK, 2011, pp. 67–71.

[7] Patrick Juola. The time course of language change. Computers and the Humanities, 2003, pp. 37(1):77–96.

[8] Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray,Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. Quantitative analysis of culture using millions of digitized books. Science, 2011, pp. 331(6014):176–182.

[9] Hyunyoung Choi and Hal Varian. Predicting the present with Google trends. Economic Record, 2012, pp. 88(s1):2–9.

[10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems, 2013, pp. 26:3111–3119.

[11] Peter Turney, Patrick Pantel, et al. From frequency to meaning: Vector space models of semantics. Journal of artificial intelligence research, 2010, pp. 37(1):141–188.

[12] Hosein Azarbonyad, Mostafa Dehghani, Kaspar Beelen, Alexandra Arkut,Maarten Marx, and Jaap Kamps. Words are malleable: Computing semantic shifts in political and media discourse.In Proceedings of the ACM on Conference on Information and Knowledge Management, Singapore , 2017, pp. 1509–1518

[13] Hao Peng, Jianxin Li, Yangqiu Song, and Yaopeng Liu. Incrementally learning the hierarchical softmax function for neural language models. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, California USA, 2017, pp. 3267–3274

[14] Yating Zhang, Adam Jatowt, Sourav Bhowmick, and Katsumi Tanaka. Omnia mutantur, nihil interit: Connecting past with present by finding corresponding terms across time. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 2015 , pp. 645–655.

[15] Steffen Eger and Alexander Mehler. On the linearity of semantic change: Investigating meaning variation via dynamic graph models. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 2016, pp. 52–58.

[16] Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. Temporal analysis of language through neural language models. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014, Baltimore, USA, pp. 61–65.

[17] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena Statistically significant detection of linguistic change. In Proceedings of the 24th International Conference on World Wide Web, 2015, Florence, Italy, pp. 625–635.

[18] Terrence Szymanski. Temporal word analogies: Identifying lexical replacement with diachronic word embeddings. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2018, Vancouver, Canada, pp. 448–453.

[19] De Saussure F. Course in general linguistics. Columbia University Press, Irvington, 2011.

[20] John Firth. A synopsis of linguistic theory, 1930-1955. Blackwell, 1957.

[21] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, Berlin, Germany, pp. 1489–1501.