Incorporating Synonym for Lexical Sememe Prediction: An Attention-Based Model

Xiaojun Kang [1], Bing Li [1], Hong Yao [1] , Qingzhong Liang [1], Shengwen Li [2,3] , Junfang Gong [2,4] and Xinchuan Li [1,*]

[1] School of Computer Science, China University of Geosciences, Wuhan 430074, China; kangxj@cug.edu.cn (X.K.); 20141002431@cug.edu.cn (B.L.); yaohong@cug.edu.cn (H.Y.); qzliang@cug.edu.cn (Q.L.)

[2] School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China; swli@cug.edu.cn (S.L.); jfgong@cug.edu.cn (J.G.)

[3] National Engineering Research Center for Geographic Information System, China University of Geosciences, Wuhan 430074, China

[4] Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen 518034, China

[*] Correspondence: lixinchuan@cug.edu.cn; Tel.: +86-27-67883716

check for
updates

Abstract: Sememe is the smallest semantic unit for describing real-world concepts, which improves the interpretability and performance of Natural Language Processing (NLP). To maintain the accuracy of the sememe description, its knowledge base needs to be continuously updated, which is time-consuming and labor-intensive. Sememes predictions can assign sememes to unlabeled words and are valuable work for automatically building and/or updating sememeknowledge bases (KBs). Existing methods are overdependent on the quality of the word embedding vectors, it remains a challenge for accurate sememe prediction. To address this problem, this study proposes a novel model to improve the performance of sememe prediction by introducing synonyms. The model scores candidate sememes from synonyms by combining distances of words in embedding vector space and derives an attention-based strategy to dynamically balance two kinds of knowledge from synonymous word set and word embedding vector. A series of experiments are performed, and the results show that the proposed model has made a significant improvement in the sememe prediction accuracy. The model provides a methodological reference for commonsense KB updating and embedding of commonsense knowledge.

Keywords: natural language processing; knowledge base; commonsense; sememe prediction; attention model

1. Introduction

In the field of Natural Language Processing (NLP), knowledge bases (KBs) play an important role in many NLP tasks. They provide rich semantic information for downstream tasks such as semantic disambiguation using WordNet's categorical information [1], bilingual embedded learning based on a multilingual KB [2]. Besides, recent researches have demonstrated that the introducing of KBs, especially commonsense KBs, not only improves the interpretability and performance of natural language processing task but also reduces the training time for machine learning [3–5].

In the commonsense KB of natural language, sememe denotes a single basic concept represented by words in Chinese and English. Linguists pointed out a long time ago that sememes are finer-grained semantic units than words [6], and a similar point is made in the theory of the universals of language [7]. For example, sememe is used as a basic representation object to reveal the relationship between

concepts and properties they possess in HowNet, which is a well-known Chinese general KB [8]. HowNet employs linguists to artificially define some 2000 sememes as the smallest, most basic units of meaning that cannot be subdivided. It uses these sememes to annotate over 100,000 words and phrases. So far, HowNet has been successfully applied to a variety of NLP tasks, such as word similarity computation [9], semantic disambiguation [10], sentiment analysis [11], and improving word vector quality [12], and evaluating the semantic reasonableness of sentences [13].

The continued emergence of new words and the semantic evolution of words, facilitated by the increase in human communication capabilities and methods, has made it necessary to frequently enrich and update KBs. However, manually labeling sememes of words is a time-consuming and labor-intensive process. Besides, it can also suffer from inconsistent annotation results if the peoples who label sememes are not a domain knowledge expert. Benefiting the semantic information contained in the pre-trained word vectors, Xie, R., et al. [14,15] proposed a series of automatic labeling methods based on word vectors. These methods model the association between word vector representations by collaborative filtering or matrix decomposition methods, achieving automatic assigning sememes into unseen words.

However, there still remain challenges in accurately predicting sememes of unlabeled words. One of the challenges is that the performance of the model is highly dependent on the quality of the pre-trained word embedding vector. Furthermore, the word vector model is relatively simple, and the word vectors it constructs may not fully represent the senses of words in the real world. Table 1 illustrates a real-world example of the word "Shen Xue (申雪)", which means "redress an injustice". In Table 1, the similar words of the word "Shen Xue (申雪)" in the word embedding vector space are more similar to skating, because there is a famous skater named "Shen Xue (申雪)". That is to say, sometimes a word's embedding vector does not capture its semantics well. This may be due to the fact that most language models learn vectors based on the assumption that co-occurrence words are similar. We argue that the obtained word-embedding vectors mainly present correlations between words, rather than similarities between words.

Table 1. Comparison of the synonyms words and the top similar words in embedding space of "Shen Xue (申雪)".

Metric	Words	Similarity in Embedding Space
Top similar words in embedding vector	Skating (滑冰)	0.617
	winter Olympics (冬奥会)	0.573
	Speedskating (速滑)	0. 536
	Ice Arena (冰场)	0. 471
	Gymnastics (体操)	0.466
Synonyms	complain of an injustice (叫屈)	0.136
	appeal for justice (申冤)	0.122
	cry out for justice (喊冤)	0.036
	exonerate (昭雪)	0.057
True sememes	Corrections (改正), result (结果), error (误)	

To address the problem, we propose to use synonyms to improve the performance of the sememe prediction. Compared to word embedding vectors, synonyms are more consistent with human cognition, thus providing more solid references for predicting sememe. More importantly, synonym acquisition does not require a lot of training like word embedding training. Assigning synonyms of words does not require specialized knowledge, thus it can be done by volunteers.

This study aims to improve the prediction accuracy of the sememes of unlabeled words by introducing synonyms. Our original contributions include: (1) By introducing synonym knowledge, a sememe prediction model is explored from the perspective of word similarities rather than word correlations. (2) An attention-based sememe prediction model that incorporates information on

synonym sets and word embeddings is developed to optimize the prediction effect through an attention strategy.

The rest of this paper is organized as follows. In Section 2, we review the related works and illustrate the limitations that remain. Section 3 details how the proposed model works. The dataset and evaluation experiments are presented in Section 4. We discuss several major factors that may affect model performance in Section 5. Section 6 concludes our work.

2. Related Work

Many KBs have been built recently for understanding the processes of NLP and improving the performance of NLP. One type of KB is known as commonsense KB, such as WordNet [16], HowNet [8], and BabelNet [17]. Compared to other types of KBs, such as Freebase [18], DBPedia [19], and YAGO [20], those manually defined commonsense KBs are richer in human knowledge and provide promising backing for various NLP tasks.

Considering that commonsense knowledge is increasing and evolving, it is important to update the commonsense KB, such as sememes of words, by automated approaches. The core of the automated process is to build intelligent algorithms that can accurately predict the sememes of unlabeled words or evolved words. To obtain higher accuracy, the algorithms may need to leverage all available knowledge.

One line of work that predicts sememes of unlabeled words was initiated using word embedded vectors. It assumes that similar words in word vector space should share the same sememes, thus sememes of unlabeled words can be inferred with pre-trained words embedded vector [21,22]. Sememe Prediction with Word Embeddings (SPWE) model first retrieves words that are similar to the vector representation of the word to be predicted and then recommends these words to the unlabeled word to be predicted, which in turn leads to the sememe prediction [14]. The paper also developed models based on matrix decomposition strategy to learn semantics and semantic relationships between words, including Sememe Prediction with Sememe Embeddings (SPSE) model and Sememe Prediction with Aggregated Sememe Embeddings (SPASE) model, and consequently predict the sememes of unlabeled words. LD-seq2seq treats sememe prediction as a weakly ordered multi-label task to label new words [23]. The models above, however, are limited by the quality of the word embedding vector, and it remains a challenge to obtain higher prediction accuracy.

To improve sememe prediction accuracy, various data have been introduced into existing prediction models. By introducing the internal structural features of words to solve the out-of-vocabulary (OOV) problem, Character-enhanced Sememe Prediction (CSP) model improves the prediction accuracy of the low-frequency words [15]. The method can alleviate the problem of large errors in the word vectors for words with fewer frequencies in the corpus. Based on the complementarity of different languages, Qi, F., et al. [24] establishes the association between semantics and cross-lingual words in the low-dimensional semantic space, and thus improves the ability of semantics prediction. Although the above work is very innovative, the employed knowledge is not very closed with sememes, and there is still a gap between the predicted results and the sememes that should be assigned.

Recently, the Sememe Prediction with Sentence Embedding and Chinese Dictionary (SPSECD) model have been proposed, which incorporates a dictionary as auxiliary information and predicts the sememe through the Recurrent Neural Network [25]. The model can account for the fact that some words have multiple senses, achieving the improvement of prediction accuracy. However, both the senses of new words and newly evolved sense of existing words cannot be presented by a dictionary in time, because it also needs time for updating. Especially, the word item in dictionaries is a very accurate expression, thus it needs more time to carefully revise new items by professional people.

3. Methodology

In our approach, we follow the basic idea of SPWE model, an assumption that similar words will share sememes. However, we argue that although word vectors can represent some semantic relatedness between words, it is not sufficient to represent the similarity of words in the real world,

and thus are limited for accurately predicting the sememes of unlabeled words. Therefore, we employ synonyms, which embed a more accurate and richer human knowledge, to achieve sememe prediction.

3.1. Score Sememes from Synonyms

In the study, words with similar semantics are grouped into the same set, which we refer to here as synonym set, $T = \{w_1, w_2, \ldots, w_i, \ldots, w_j, \ldots, w_n\}$, where w_i denotes a word. Any two words, w_i and w_j, in the same synonym set are synonymous.

A score function is defined to score all the candidate sememes of unlabeled word w, in which high-scored sememes will be predicted as the sememes of w. For incorporating the knowledge in pre-trained word vectors, the distance of words in the pre-trained vector space is employed in the function. The function, using synonyms, can be formulated as Equation (1):

$$Score_{SPS}(s_j, w) = \sum_{w_i \in T} \cos(w, w_i) \cdot M_{ij} \cdot c^{r_i} \tag{1}$$

where M is the matrix representing the relationship between words and sememes, and can be calculated as

$$M_{ij} = \begin{cases} 1 & s_j \text{ is sememe of } w_i \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

and $cos(w, w_i)$ presents the cosine distance between the embedding vector of w and that of w_i. Different from the classic collaborative filtering in recommendation systems, sememes of most unrelated words do not include the true sememes of w in sememe prediction task. Therefore, the score function should give significantly large weight to the most similar words. To increase the influences of a few top words that are similar to w, a declined confidence factor c^{r_i} for each word w_i is assigned, where r_i the similarity rank of the word is w_i. to the word w in embedding space.

3.2. Attention-Based Sememe Prediction

Although synonyms can more accurately depict semantic similarity between two words than word embedding vector, the number of words existing in the synonym dataset is far fewer than the number of words represented in the pre-trained word vector dataset, such as Glove [26]. For words that are not included in synonym datasets, the above score function does not yet fully support the task of sememe prediction. Besides, prediction accuracy may also be impaired for words with fewer synonyms. Therefore, we combine synonym sets and pre-trained word vectors to depict the semantic similarity between words. A straightforward model can be derived, which score recommendation sememes by summing the scores of the two models using a coefficient of weight, as shown in Equation (3).

$$Score_{SPSW}(s_j, w) = \alpha Score_{SPS}(s_j, w) + (1 - \alpha) Score_{SPWE}(s_j, w) \tag{3}$$

where α is a hyperparameter, which denotes the weight of the SPS model's score.

Actually, we found that the predicted sememes based on synonym and based on word vectors, such as SPWE, were significantly different for different words. Using Equation (2) weights presented by the hyperparameter α is relatively straightforward, it is not flexible enough to make full use of knowledge from both the synonym and word embedding.

The study assumes that the weights of different knowledge should vary for different unlabeled words that are to be predicted. Inspired by [27], this study introduces an attention mechanism to obtain those weights. One of the benefits of attention mechanisms is that they allow for dealing with variable inputs, focusing on the most relevant parts of the input to make decisions [28]. An attention function can be described as mapping a query and a set of key-value pairs to an output [27], where the query and keys are word vectors; output is the weights of related words. Thus, an attention-based model,

named ASPSW (Attention-based Sememe Prediction combining Synonym and Word embedding), is derived, and its score function can be calculated as Equation (4):

$$Score_{ASPSW}(s_j, w) = Attn(s_j, w, W) = \sum_{W_i \in Y}^{n} a_i^{Attn} Score(s_j, w) \tag{4}$$

where a_i^{Attn} denotes the weights of contributions to different knowledge for different sememes in the joint model. The difference can be adjusted according to the distance in the word embedding space. Based on this, the weights of the contributions of different knowledge can be calculated by dynamically adjusting the score weights of the knowledge from Synonym and pre-trained word vector:

$$a_i^{Attn} = \begin{cases} \dfrac{1}{2 - log(|Sim_{we} - Sim_{sy}|)} & if\ w_i \in W \\ \dfrac{1 - log(|Sim_{we} - Sim_{sy}|)}{2 - log(|Sim_{we} - Sim_{sy}|)} & if\ w_i \in T \end{cases} \tag{5}$$

$$Sim_{we} = \frac{1}{K} \sum_{w_i \in W} cos(w,\ w_i) \tag{6}$$

$$Sim_{sy} = \frac{1}{N} \sum_{w_i \in T} cos(w,\ w_i) \tag{7}$$

where T is the synonym set of word w; W presents the top K similar words set of w in embedding space respectively, where K is a hyperparameter; Sim_{we} and Sim_{sy} represent the average semantic similarity between new words and similar words in word embedding and synonyms, respectively; $cos(w, w_i)$ is the cosine similarity between w and w_i according to their embedding vectors.

4. Experiment and Results

4.1. Dataset

HowNet: HowNet is a commonsense KB, in which approximately 2,000 sememes are manually defined. Those sememes serve as the smallest unit of meaning that is not easily re-divided, and more than 100,000 words and phrases are annotated with these sememes. The structure of HowNet is illustrated in Figure 1. The example in the figure shows the word "草根" explained in terms of sememes. The word consists of two senses in Chinese. One is "Grass root", which means a certain organ of a plant, and the other is "Grass roots", which generally refers to people at the bottom level or entrepreneurs starting from scratch. The former is explained by sememes, "part", "base" and "flowerGrass", and the latter consists of sememes, "human" and "ordinary". To reduce the noises from low-frequency sememes, the study removed the low-frequency sememes following the approach in [14] and experimented with only 1,400 remaining sememes.
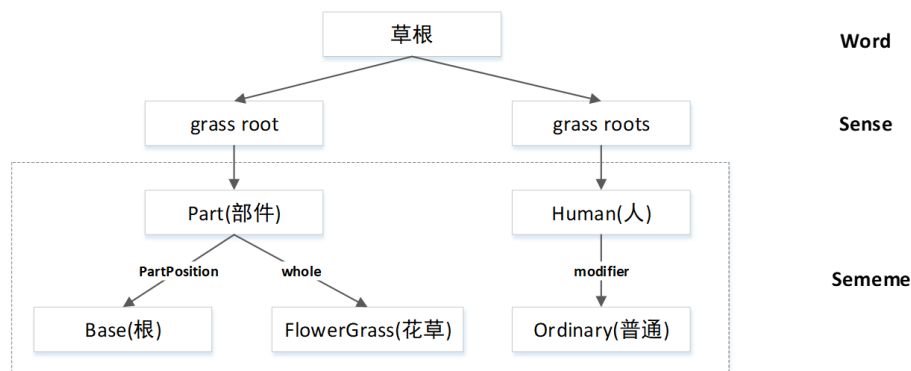

Figure 1. The sememes of Word "草根" in a commonsense database.

Sogou-T: The Sogo-T Corpus is an Internet corpus developed by Sogou and its corporate partners, which contains a variety of original web pages from the Internet, with a total of about 2.7 billion words.

Synonym dictionary: There are several available synonym data sources, such as the synonym dictionary ABC Thesaurus, the Chinese Dictionary, HIT IR-Lab Tongyici Cilin from Harbin Institute of Technology Social Computing and Information Retrieval Research Center, China. In the experiment, we selected HIT IR-Lab Tongyici Cilin (Extended) as a data source of the synonym set. It contains a total of 77,343 words. All words are organized together in a tree-like hierarchy with a total of five layers, as shown in Figure 2. For each layer, each category corresponds to a different code, e.g., "Evidence", "Proof" belong to the same category with code "Db03A01". The lower the layer, the finer the granularity of the category and the more similar the sense of words under the same node. The study uses only the lowest layer to construct synonym sets.
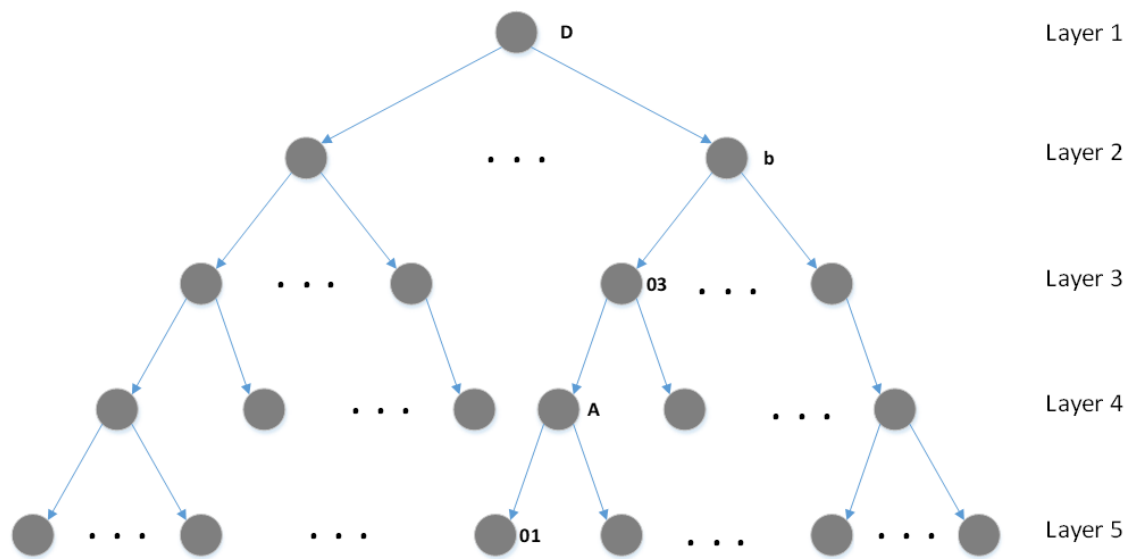

Figure 2. The five layers in HIT IR-Lab Tongyici Cilin.

4.2. Experimental Settings

The study employs Glove [26] to obtain the word embedding vectors of all words in the Sogou-T corpus. To keep data alignment, we removed words that were not contained in the pre-trained word vectors or not listed in the synonym sets. In the end, we selected a total of 44,556 words from HowNet. Ten percent of the words are selected for the test, and the rest 90% words are for training.

Models in three state-of-the-art works are selected as baseline models. The first work [14] includes five models, SPWE, SPSE, SPASE, SPWE+SPSE, and SPWE+SPASE. The second group models are proposed in [15], including five models variants: SPWCF (Sememe Prediction with Word-to-Character Filtering); SPCSE (Sememe Prediction with Character and Sememe Embeddings); SPWCF+SPCSE models that use only the internal information of words and both internal and external information for the original meaning; and the integrated framework of prediction CSP (Character-enhanced Sememe Prediction), respectively. The model in the last group is LD-seq2seq (Label Distributed seq2seq) that treats the sememe prediction as a weakly ordered multi-label task [23].

Follow the settings in [14], all the dimension sizes of word vectors, sememe vectors, and character vectors are set to 200. For the baseline model, in the SPWE model, the hyperparameter c that controls the contribution weight of different words is set to 0.8. The number of semantically similar words in the word vector space is set to $K=100$, which is the same as the setting in work [14]. In the SPSE model, the probability of decomposing zero elements in the matrix of word-sememe is set to 0.5%, the initial learning rate is set to 0.01, and the learning rate drops after iteration, and $\lambda_{SPWE}/\lambda_{SPSE}$ is set to 2.1 in its joint model, where λ_{SPWE} and λ_{SPSE} represent the weights of the SPWE and SPSE models, respectively.

For models from [15], we use cluster-based character embedding [29] to learn pre-trained character embeddings; the probability of decomposing zero elements in the matrix of word-sememe is set to 2.5%. For the joint model, we set the weight ratio of SPWCF and SPCSE to 4.0, the r weight ratio of SPWE and SPSE is 0.3125, and the weight ratio of internal and external models is 1.0. For LD-seq2seq [23] model, the dimension size of all hidden layers is set to 300, and its training batch size is set to 20. For SPSW model, we argue the contributions from SPS and SPWE are approximately equivalent, so α is set to 0.5.

4.3. Results

Since a large number of words have multiple sememes, the sememe prediction task can be considered as a multi-label classification task. The study uses the Mean Average Precision (MAP) as a metric, which is the same as previous work [14], to evaluate the accuracy of predicting sememe. For each unlabeled word in the test set, our model and the baseline models ranked all candidate sememes. Their MAPs are calculated by ranked results on the test dataset and are reported in Table 2.

Table 2. Prediction accuracy: Mean Average Precision (MAP); the best result is in bold-faced.

Model	MAP
SPWE [14]	0.5610
SPSE [14]	0.3916
SPASE [14]	0.3538
SPWE+SPSE [14]	0.5690
SPWE+SPASE [14]	0.5684
SPCSE [15]	0.3105
SPWCF [15]	0.4529
SPWCF+SPCSE [15]	0.4849
CSP [15]	0.6408
LD-seq2seq [23]	0.3765
SPS	0.5818
SPSW	0.6578
ASPSW	**0.6774**

Table 2 shows the accuracy of the sememe prediction accuracies of the baselines model and the proposed models in the study, where SPS, SPSW, and ASPSW are three models that employ Equation (1), (2), and (3) as score function, respectively.

The results suggest the proposed models ASPSW had made significant improvements compared to SPWE model. This experimental result further supports our idea that synonym sets, compared to word vectors, can more accurately characterize the sememe correlated relationships between words. The SPSW model has a larger gain than the SPS model, which shows that although the synonymy forest can provide more accurate semantic similarity, the synonyms provided by the synonymy forest are limited and rare, so the semantic information provided by the word vector can be combined to further improve the accuracy of the prediction of sememes. The ASPSW, using attention strategy to dynamic weigh model significantly, outperforms the fixed weights, which shows that the proposed attention mechanism is effective in predicting the semantics for different unlabeled words and can effectively adjust the effects of different knowledge for words to be predicted.

5. Discussion

5.1. The Two Ways of Combining Synonyms and Word Embedding Vectors

Two score functions are introduced in Section 3.2 for combining knowledge from synonyms and word embedding vector. One is the static SPSW, as shown in Equation (2), and the other is attention-based ASPSW, as shown in Equation (3). The former score function combines the knowledge between synonyms and from pre-trained word vector by the hyperparameters, α, and the later score

function dynamically balances two kinds of knowledge using an attention strategy. To examine the performance of two models, we performed experiments with a different value on static SPSW, and listed the results in Table 3.

Table 3. The prediction accuracy of different α; the best result is in bold-face.

α	MAP
0.1	0.5820
0.2	0.6023
0.3	0.6222
0.4	0.6416
0.5	0.6578
0.6	0.6718
0.7	**0.6787**
0.8	0.6764
0.9	0.6674
ASPSW	0.6774

As shown in Table 3, the values of α have made a significant effect on the prediction accuracy. When it was set to 0.7, the model SPSW achieved the best results, and the ASPSW obtained the second-best results. Despite an appropriately selected α value, static SPSW achieves better results, the best and the second results are a little different. Considering the robustness of methods, we argue that ASPSW is a more promising model for sememe prediction.

To observe the difference caused by models, we performed experiments on random-selected 100 words with three typical models (score function), SPWE, SPS, and ASPSW. The scores of the three models are recorded and plotted in Figure 3. The figure shows that some of the scores of the SPS model are close to 0, which may be because the knowledge in the synonym dictionary is incomplete. For a new word, SPS can rarely find a valid synonym for inferring sememes. In most cases, the prediction score of the ASPSW model is higher than that of the SPWE model and the SPS model, indicating that the dynamical weights in the joint model can make full use of different knowledge and avoid false predictions due to incompleteness or inaccuracy in a single type of knowledge.
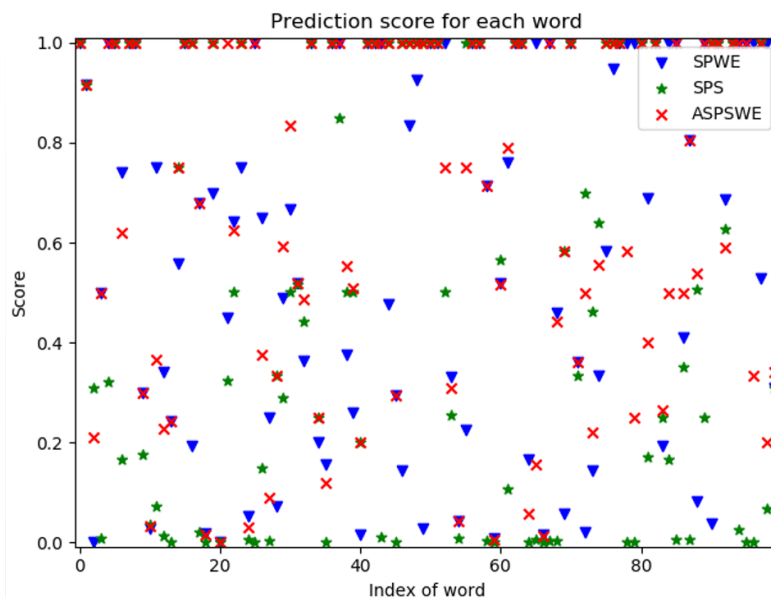

Figure 3. Randomly select 100 words and test their sememes prediction score (MAP value) of the SPWE, SPS, and ASPSW models.

5.2. Impact of the Value of K

The parameter *K* is the number of similar words in the word vector space used to select candidate sememe. As a hyperparameter, the size of *K* may affect the prediction accuracy of the proposed model. To examine the effect of the value of *K*, we set the value *K* from 10 to 100. The accuracies of SPWE and the proposed model, ASPSW, are listed in Table 4.

Table 4. Prediction accuracy of the SPWE and ASPSW models under setting different values of k;the best results are in bold-face.

Nearest Word Number	SPWE	ASPSW
10	0.5478	0.6724
20	0.5566	0.6762
30	0.5587	0.6773
40	0.5597	**0.6778**
50	0.5602	**0.6778**
60	0.5605	0.6776
70	0.5606	0.6775
80	0.5608	0.6777
90	0.5609	0.6777
100	0.5610	0.6774

As shown in Table 2, ASPSW provides great prediction accuracy; even the value of K is set to small values. When *K* is set to larger than 20, the prediction results tend to be stable, indicating that the model has good robustness. From this, it suggests that smaller numbers of the most similar words will cover the semantics of the words, thus achieving a quite accurate prediction of sememes. The results further confirm that in Table 1, although the synonym KB provides a few synonyms, it is still possible to reach an accuracy that exceeds the baseline. As the values of *K* increase, the prediction accuracy of the model improves. In the process, the prediction accuracy of the ASPSW model is kept well above the accuracy of the baseline model, SPWE, demonstrating the validity of the ASPSW model.

5.3. Calculation Performance Analysis

In the experiment, we examined the time efficiency of different models on predicting the sememes of unlabeled words. As shown in Table 5, we randomly selected 5000 words as a test task for predicting their sememes with different models and recorded the time consumption of the training process and prediction, respectively.

Table 5. Time consumption for predicting sememes of 5000 unlabeled words.

Method	Training Costs (s)	Predicting Costs(s)	Total (s)
SPWE	NA	2129	2129
SPSE	6510	40	6550
SPWE+SPSE	6510	2195	8705
SPCSE	41,191	2031	43,222
SPWCF	NA	334	334
SPWCF+SPCSE	41,191	2417	43,608
CSP	47,701	4656	52,357
SPS	NA	22	22
SPSW	NA	2169	2169
ASPSW	NA	2639	2639

Table 5 shows that the SPSY model takes the least amount of time to accomplish this task. It benefited from the fact that the model does not contain training the process of the reference words, synonym, and thus, it does not need to calculate word similarities with word vectors. Actually, all the

models without the training process spend less time than the models that contain a training process, because the training process is very time-consuming. Although the SPS model based on matrix decomposition and SPWCF model based on internal character features of words can complete the prediction process in a relatively short time, their prediction accuracy still remains lower.

In addition, compared with SPSE and SPCSE models, the SPSW and ASPSW model does not require additional time for training. The SPSW model based on fixed weights is similar to the SPWEA model based on word embedding in time consumption. The ASPSW model based on an attention mechanism can also improve the prediction accuracy of sememes without significantly increasing time consumption.

5.4. Case Study

In the case study, we give further analysis by detailed examples to explain the effectiveness of our model.

Table 6 lists the results of some of the SPWE model and ASPSW model sememe predictions. Each word shows its top five predicted sememes, in which the true sememes are in bold. As it can be seen from the table, ASPSW predicts the true sememes in their top positions, thus showing that the finding of semantically similar words is crucial for the sememe prediction of words. In the SPWE model using the word vector only, the corrected predicted sememe of words such as "saber" and "pull, social connections" do not rank in top positions. For the word "saber", the vector-based model focuses more on the semantics of the simultaneous occurrence of the word "knife", so that sememe "tools" and "cutting" rank higher than the correct sememe "army" and "weapon". With the introduction of the synonym set, the ASPSW model can compensate for the inability of word embedding to accurately define semantics and make the recommended sememe for "saber" more biased towards the sememes of "army" and "weapon". In addition, for words such as "appease" and "old woman", the SPWE model failed to predict correct sememes. For example, the SPWE model does not capture the semantic information of the word "appease", and the recommended sememe is all semantics that is not closed to "appease". The introduction of the ASPSW model with a synonym set achieves good prediction results, which further demonstrates that word embedding has a significant gap in the capture of semantic information from the synonym set.

Table 6. Comparison of sememe prediction examples for the SPWE and ASPSW models, the sememes in bold font are the true sememes for each word.

Words	Top 5 Sememes with SPWE	Top 5 Sememes with ASPSW	True Sememes
Saber (军刀)	tools, Cutting, Breaking, **Army**, **Weapons** (用具, 切削, 破开, **军**, **武器**)	**army**, **weapons**, tools, cutting, **piercing** (**军**, **武器**, 用具, 切削, **扎**)	Army, Weapons, Piercing (军, 武器, 扎)
Kindergarten (幼儿园)	**place**, education, teaching, learning, **people** (**场所**, 教育, 教, 学习, **人**)	**place**, **people**, **children**, **care**, education, people (**场所**, **人**, **少儿**, **照料**, 教育)	people, place, children, care (人, 场所, 少儿, 照料)
special column (专栏)	Chinese, **books**, publishing, news, time (语文, **书刊**, 出版, 新闻, 时间)	**book**, **special**, Chinese, publishing, news (**书刊**, **特别**, 语文, 出版, 新闻)	Parts, Books, special (部件, 书刊, 特别)
appease (息怒)	person, be kind, answer, sit, emperor (人, 善待, 答, 坐蹲, 皇)	**emotion**, **angry**, **stop**, person, be kind (**情感**, **生气**, **制止**, 人, 善待)	emotion, angry, stop (情感, 生气, 制止)
pull, social connections (门路)	rich, become, **method**, **person**, **intimate** (富, 成为, **方法**, **人**, **亲疏**)	**method**, **person**, **intimate**, success, road (**方法**, **人**, **亲疏**, 成功, 道路)	person, method, intimate (人, 方法, 亲疏)
old woman (妪)	crying, poultry, shouting, diligent, surname (哭泣, 禽, 喊, 勤, 姓)	**person**, **elderly**, **female**, crying, poultry (**人**, **老年**, **女**, 哭泣, 禽)	person, elderly, female (人, 女, 老年)

To better illustrate the difference effect over different words, we took two more words as an example, and distinguish their similar words by whether they contain correct sememes in the pre-trained word vector space. As shown in Figure 4a, the top similar words to word "申雪" in the vector space do not contain the sememe that should recommend the word "申雪". For the unlabeled word "便士", as shown in Figure 4b, the words which contain the same sememe with it are clustered around it in the vector space. The two examples show that there is a very clear deviation in the distribution of similar words in word vector; this may be caused by the fact that the language model of generating word embedding vectors is inferred from word co-occurrence instead of similar semantics. To overcome those deviations, we suggest again that it is very necessary to combine the synonym and pre-trained word vector for better understanding word embedding vectors and improving the performances of various downstream tasks.
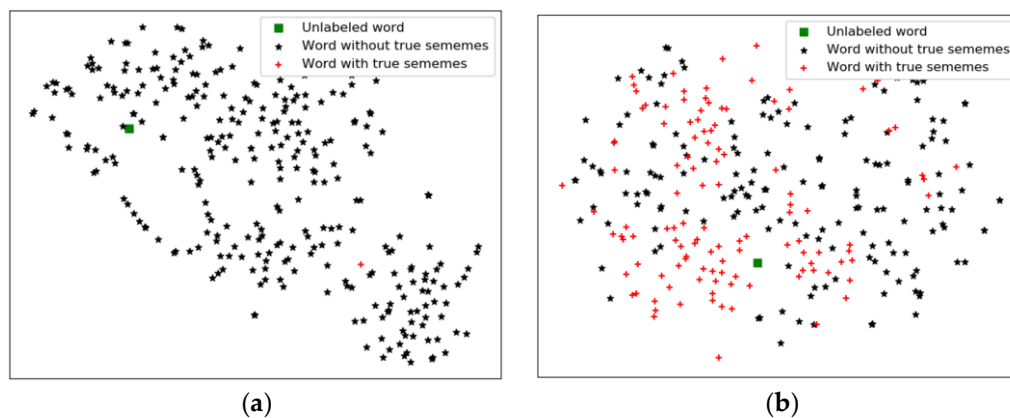

(**a**)   (**b**)

Figure 4. Top 300 similar words of the unlabeled word (**a**) "申雪" (**b**) "便士" in the word vector space, where "+" means that the sememes can be recommended for the unlabeled word because the word contains the true sememe of the unlabeled word; "*" presents a word that it is impossible to recommend the true sememe for unlabeled word because the word and the unmarked word do not contain the same sememes.

6. Conclusions and Future Work

In this study, we propose to predict the sememes of unlabeled words by introducing a synonym. An attention-based model, ASPSW, is developed that incorporates similar relationships in the synonym set into the sememe prediction decisions. A series of experiments are performed, and the results show that the proposed model has made a significant improvement in the sememe prediction accuracy. This study suggests that the dynamical fusion of knowledge from different sources is expected to enhance the ability to perform NLP tasks, especially in the absence of training samples.

In our future work, we will make the following efforts: (1) There is a tree-like hierarchy structure in HowNet dataset, and we plan to merge the hierarchical relationships between the sememes into future prediction models, which may improve the accuracy of sememe prediction; (2) more synonym datasets, including WordNet, will be combined to improve the performance of sememe prediction.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aouicha, M.B.; Taieb, M.A.H.; Marai, H.I. WordNet and Wiktionary-Based Approach for Word Sense Disambiguation. In *Transactions on Computational Collective Intelligence XXIX*; Springer: Cham, Switzerland, 2018; pp. 123–143.
2. Artetxe, M.; Labaka, G.; Agirre, E. Learning bilingual word embeddings with (almost) no bilingual data. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2017; Volume 1, pp. 451–462.
3. Chen, Y.; Luo, Z. A Word Representation Method Based on Hownet. *Beijing Da Xue Xue Bao* **2019**, *55*, 22–28.
4. Peng-Hsuan, L. CA-EHN: Commonsense Word Analogy from E-HowNet. *arXiv* **2019**, arXiv:1908.07218.
5. Iqbal, F.; Fung, B.C.M.; Debbabi, M.; Batool, R.; Marrington, A. Wordnet-based criminal networks mining for cybercrime investigation. *IEEE Access* **2019**, *7*, 22740–22755. [CrossRef]
6. Bloomfield, L. A set of postulates for the science of language. *Language* **1926**, *2*, 153–164. [CrossRef]
7. Goddard, C.; Wierzbicka, A. *Semantic and Lexical Universals: Theory and Empirical Findings*; John Benjamins Publishing: Amsterdam, The Netherlands, 1994; Volume 25.
8. Dong, Z.; Dong, Q. *Hownet and the Computation of Meaning*; World Scientific: Singapore, 2006; pp. 1–303.
9. Liu, Q.; Li, S. Word similarity computing based on Hownet. *Comput. Linguist. Chin. Lang. Process.* **2002**, *7*, 59–76.
10. Duan, X.; Zhao, J.; Xu, B. Word sense disambiguation through sememe labeling. In Proceedings of the International Joint Conference on Artificial Intelligence, Hyderabad, India, 6–12 January 2007; pp. 1594–1599.
11. Huang, M.; Ye, B.; Wang, Y.; Chen, H.; Cheng, J.; Zhu, X. New word detection for sentiment analysis. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, USA, 22–27 June 2014; Volume 1, pp. 531–541.
12. Yang, L.; Kong, C.; Chen, Y.; Liu, Y.; Fan, Q.; Yang, E. Incorporating Sememes into Chinese Definition Modeling. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *28*, 1669–1677. [CrossRef]
13. Liu, S.; Xu, J.; Ren, X. Evaluating semantic rationality of a sentence: A sememe-word-matching neural network based on hownet. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Dunhuang, China, 9–14 October 2019; pp. 787–800.
14. Xie, R.; Yuan, X.; Liu, Z.; Sun, M. Lexical sememe prediction via word embeddings and matrix factorization. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 4200–4206.
15. Jin, H.; Zhu, H.; Liu, Z.; Xie, R.; Sun, M.; Lin, F.; Lin, L. Incorporating Chinese Characters of Words for Lexical Sememe Prediction. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; Volume 1.
16. Miller, G.A. WordNet: A Lexical Database for English. *Commun. ACM* **1995**, *38*, 39–41. [CrossRef]
17. Navigli, R.; Ponzetto, S.P. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.* **2012**, *193*, 217–250. [CrossRef]
18. Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; Taylor, J. Freebase: A collaboratively created graph database for structuring human knowledge. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Vancouver, BC, Canada, 10–12 June 2008; pp. 1247–1249.
19. Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. DBpedia: A nucleus for a Web of open data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4825 LNCS, pp. 722–735.
20. Hoffart, J.; Suchanek, F.M.; Berberich, K.; Weikum, G. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artif. Intell.* **2013**, *194*, 28–61. [CrossRef]
21. Rizkallah, S.; Atiya, A.F.; Shaheen, S. A Polarity Capturing Sphere for Word to Vector Representation. *Appl. Sci.* **2020**, *10*, 4386. [CrossRef]
22. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the NAACL-HLT, Minneapolis, MN, USA, 2–7 June 2019.
23. Li, W.; Ren, X.; Dai, D.; Wu, Y.; Wang, H.; Sun, X. Sememe prediction: Learning semantic knowledge from unstructured textual wiki descriptions. *arXiv* **2018**, arXiv:1808.05437.

24. Qi, F.; Lin, Y.; Sun, M.; Zhu, H.; Xie, R.; Liu, Z. Cross-lingual Lexical Sememe Prediction. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October– 4 November 2018; pp. 358–368.

25. Bai, M.; Lv, P.; Long, X. Lexical Sememe Prediction with RNN and Modern Chinese Dictionary. In Proceedings of the 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Huangshan, China, 28–30 July 2018; pp. 825–830.

26. Pennington, J.; Socher, R.; Manning, C.D. GloVe: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.

27. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.

28. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018-Conference Track, Vancouver, BC, Canada, 30 April–3 May 2018.

29. Chen, X.; Xu, L.; Liu, Z.; Sun, M.; Luan, H. Joint learning of character and word embeddings. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.