Semantic Capture Analysis in Word Embedding Vectors Using Convolutional Neural Network

Raúl Navarro-Almanza[(✉)], Guillermo Licea[(✉)], Reyes Juárez-Ramírez[(✉)], and Olivia Mendoza[(✉)]

Universidad Autónoma de Baja California, Calzada Universidad 14418, Tijuana, Baja California, Mexico
{rnavarro,glicea,reyesjua,omendoza}@uabc.edu.mx
http://fcqi.tij.uabc.mx/

Abstract. The semantic relation detection among entities from unstructured text is an important task in automatic knowledge construction to discover new knowledge. Word embeddings have been successful in capturing semantic relations among entities in unstructured text. In this work we propose to use WordNet as a knowledge base to extract semantic relations among entities and measure how well word embeddings vectors capture semantic regularities by themselves, using state-of-art classification model to detect semantic relations. We present semantic relation capture f-measure score in word embedding vectors of 94.9%, the semantic relations addressed in this work are taxonomic relations (*hypernym-hyponym*) and part-of relations (*holonym-meronym*).

Keywords: Semantic regularities · Word embedding · Relation classification · Convolutional neural network

1 Introduction

The relation detection among entities is an important task in automatic knowledge base construction (KBC), brings capability of reasoning and discover new knowledge from existing one.

There is special attention to relation detection without handcrafted features, overcoming the dependency and limited application domain that implies those approaches. Handcrafted resources such as WordNet[1] are used on generic domains [2,4,15,18], however, on more specialized domains are more difficult to find structured data to improve KBC. Through more linguistic independent and automatic methods to capture semantics improve automatic KBC processes on specialized domains, especially when knowledge bases are constructed from unstructured data.

One of the novel word representation models is neural word embeddings, which is the basis of our proposal. Embedding word representation aims to reduce a word vector representation into lower dimensionality and continuous vector space.

---

[1] https://wordnet.princeton.edu/wordnet/.

Neural network language models have been used on word embeddings [1,12,13]. In [12], it is demonstrated that many syntactic and semantic regularities can be captured in those embedding representations.

The available datasets for relation classification such as SemEval 2010 Task 8 [8] promote comparison between state of the art results and naturally thrust to make more prominent models [5,11,14,16]. In this dataset there is various kind of relations, however, in this work we are interested in classifying more basic relations such as *hypernym-hyponym* and *holonym-meronym*, because of that, we use an external knowledge base to evaluate those kinds of relations.

The aim of our work is to measure how well word embeddings capture semantic relations among words. The actual approaches to classify semantic relations such as *cause-effect*, *producer-product*, *message-topic*, *content-container*, etc. use a context set of words along with entities to predict the class of semantic relation [11,14]. However, we propose to learn more basic semantic relations by using only word embedding representation without context set of words.

The classifier that we choose for the assessment is state of the art in semantic relation detection [14]. There is a consensus that Convolutional Neural Networks (CNN) assess good results. In this work those models are adapted to learn only two-word embedding vectors as input to classify their semantic relation.

The remainder of the paper is structured as follows. Section 2 contains related works with word embedding to semantic relation classification. Section 3 describes the methodology that guides this study. Section 4 details the parameters of the experiment setup. Section 5 presents the result of the experiment and discussion. Section 7 presents future work.

2   Related Work

The use continuous vector space representation to identify syntactic/semantic relations between words are increasing recently [5–7,9,17] because of their generality to capture syntactic/semantic regularities from unstructured text data.

There are studies where rely on the semantic captured on the continuous vector space representations to detect semantic relations. In [6], it is observed that there are non-linear regularities between vectors that capture semantic relations, for example, the well-known expression $v(\texttt{king}) - v(\texttt{queen}) \equiv v(\texttt{man}) - v(\texttt{woman})$ does not represent the *hypernym-hyponym* relation over all vector space, that is to say, we can not generalize that the vector $v(\texttt{man}) - v(\texttt{king})$ represents *hypernym-hyponym* relation and infer new hierarchies in all vector space. They analyzed that there are regions where those regularities are locally shared between words, one of our hypothesis is that those regions can be learned by a non-linear classifier, due to CNN are the state-of-art models to learn relations regularities among words represented on continuous vector space [3,14,16]. We selected those models to learn patterns of local regions where semantic regularities are shared.

3   Methodology

The methodology to implement our proposal on semantic relation extraction is shown in the Fig. 1. The data preparation task consists on four steps: (1) Word embedding: word embedding are generated using skip-gram model [13], trained with Wikipedias[2] articles; (2) Entities selection: the terms (or entities) are selected from WordNet if they exist on word embedding dictionary (obtained on previous step); (3) Relation extraction: in this step the semantic relations *hypernym*, *hyponym*, *holonym* and *meronym* between selected terms are extracted from WordNet; (4) Word embedding representation: the terms extracted in the 2nd step are converted to continuous vector space representation.
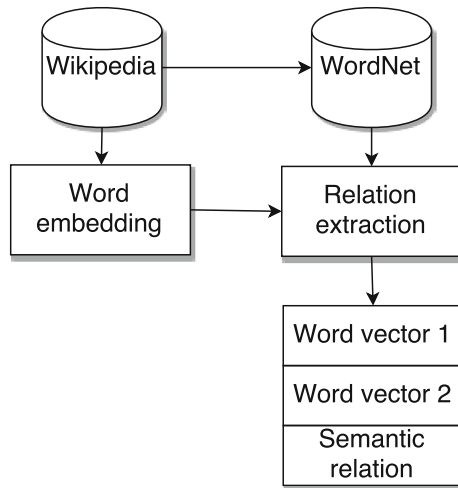
Fig. 1. Semantic relations extraction methodology

3.1   Word Embeddings

The first step in our process the word embeddings is generated using skip-gram model proposed by [13] trained on Wikipedia articles (dump 2016). Skip-gram model aims to predict context words $w_{context} = (w_{t-i}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+i})$ where $i \geq 1$ given a word $w(t)$ as input (Fig. 2). The projection is a low-dimensionality and continuous vector space representation of the word $w(t)$.

3.2   Entities Selection

The entities used to get semantic relation between them must satisfy the condition that both entities (e_i and e_2) given a relation $\mathcal{R}_i = (e_1, e_2, s_i)$ have to
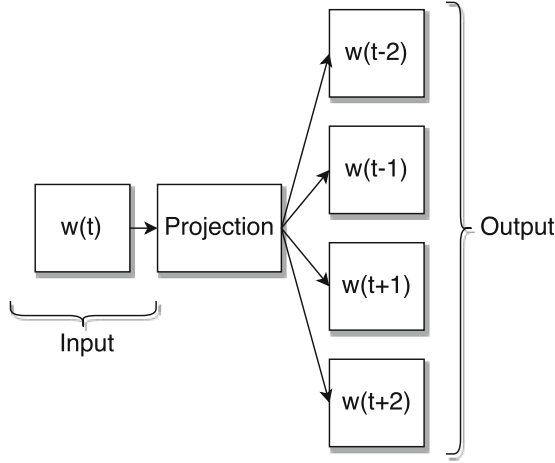

Fig. 2. Skip-gram model architecture [13]

exist in word embedding space, where s_i represents a semantic relation type. In other words, both entities must exist on the corpora used to train neural network language model.

3.3 Semantic Relation Extraction

The semantic relations extracted from WordNet are represented by te set $\mathcal{S} = \{hypernym,\ hyponym,\ holonym,\ meronym\}$. The semantic relations are extracted by follows. For each entity (e_i) that exists in the selected entities from previous step, the entities that are related (by $\mathcal{S}$ relations) and exists in the selected entities, form a relation $\mathcal{R}_i = (e_1, e_2, s_i)$, where $s_i \in \mathcal{S}$. In the Table 1 are shown equivalence relations between selected semantic relations, in Sect. 5 discuss how those equivalences can be use to empirically evaluate classifier result.

Table 1. Semantic relations

Relations	Representation	Inverse relation
Hypernym	$X \overrightarrow{hypernym} Y$	$Y \overrightarrow{hyponym} X$
Hyponym	$X \overrightarrow{hyponym} Y$	$Y \overrightarrow{hypernym} X$
Holonym	$X \overrightarrow{holonym} Y$	$Y \overrightarrow{meronym} X$
Meronym	$X \overrightarrow{meronym} Y$	$Y \overrightarrow{holonym} X$

3.4 Semantic Relation to Word Embedding Representation

Each one of relations extracted in the previous step is represented as matrix form $(2x\mathcal{D})$, where $\mathcal{D}$ are the dimension of the vector space representation. All entities

on the relations are in continuous vector space and associated to the category of relation class ($\mathcal{S}$).

3.5   Evaluation

To measure how well word embedding capt the semantic relation in the corpus, we use accuracy (Eq. 1), precision (Eq. 2), recall (Eq. 3) and f-measure (Eq. 4) metrics, and cross validation k-Fold, where k is 8.

Accuracy

$$Accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \hat{y}_i = y_i \tag{1}$$

where, y_i is the target class and $\hat{y}_i$ output class.

Precision

$$Precision = \frac{tp}{tp + fp} \tag{2}$$

where, tp are the true positives and fp false positives.

Recall

$$Recall = \frac{tp}{tp + fn} \tag{3}$$

where, tp are the true positives and fn false negatives.

F-measure

$$Fmeasure = \frac{2 * precision * recall}{precision + recall} \tag{4}$$

4   Experiment Setup

4.1   Word Embedding

The classifier's input is a matrix with size 2×400 formed by row vectors which are from `word2vec` trained with Wikipedia corpus using skip-gram with window size value of 5 and embedding dimension of 400.

4.2   Semantic Relations

WordNet is used as knowledge base to extract semantic relations among words that exists on word embedding space. Randomly $9,000$ relations are selected for each class ($hypernym, hyponym, holonym, meronym$) to maintain balanced number of instances (36,000 instances in entire dataset) (Fig. 3).
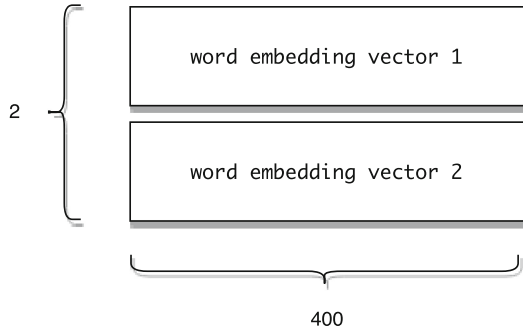
Fig. 3. Input shape

4.3   Convolutional Neural Network

The classifier's input is a concatenation of word embedding vectors using skip-gram model. Based on the state of art models for semantic relation classification, we use a CNN to learn semantic relation regularities over word vectors. The overview of the architecture used in this experiment is shown in the Fig. 4. This architecture is a simplified version of the state-of-art models to semantic relation classification [14], the aim of this work is measure the semantic capture by word embedding, thus the context word set are removed from our model, and the architecture are simplified.

In the Table 2 are shown the configuration of each layer used in the proposed CNN. Due to input dimension, we set filter size of 2×2. In the fully connected
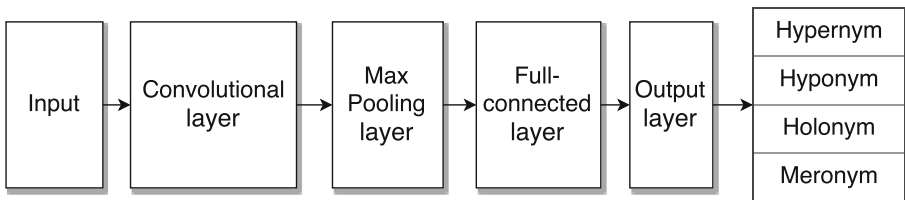

Fig. 4. CNN architecture

Table 2. CNN layers dimension

Layer	Dimension
Input	2×400
Convolution	100
Pooling	100
Full-connected	800
Output	4

layer has ReLU activation function and a Dropout of 0.4. The learning algorithm selected to this problem was Adam [10] with learning rate $\lambda = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e - 08$. The number of epochs to train the CNN was set to 7.

5   Result and Discuss

The mean of K-Folds validation where $k = 8$ on the F-measure score achieved 0.946. The results by fold are shown in the Table 3, they are the mean of validation by epoch on each fold. As is shown in other works, word embedding achieves good results on semantic capture, and there are used to semantic relation detection. However, these models are developed using context word vectors to infer various semantic regularities between words.

Taking into consideration that only embedding word representation was conducted, the result of evaluation of semantic capture in word embedding show good results (achieving F-measure score of **94.9%**) on *hypernym, hyponym, holonym* and *meronym* relations detection.

The transitivity in the relations (as shown in Table 1) can be used to validate the result of the classifier, given a tuple $\mathcal{T} =< entity1, entity2 >$ and result relation $\mathcal{R}_i$, it can be and empirically evaluation that the classifier achieve a good result if changing the entities order ($< entity2, entity1 >$) the result is semantically inverse to relation $\mathcal{R}_i$, eg. $< dog, canine >$: *hyponym* $\equiv < canine, dog >$: *hypernym*.

Table 3. Experiment results

Fold	Accuracy	Precision	Recall	F-measure
1	0.943	0.952	0.943	0.948
2	0.934	0.940	0.933	0.937
3	0.940	0.947	0.943	0.945
4	0.946	0.953	0.943	0.948
5	0.950	0.957	0.947	0.952
6	0.945	0.951	0.950	0.950
7	0.951	0.953	0.950	0.952
8	0.959	0.962	0.961	0.961
Average	0.946	0.952	0.946	0.949
σ	0.0073	0.0064	0.0078	0.0069

6   Conclusions

In this work a semantic capture analysis was conducted on word embedding vectors. In the evaluation process, we compare the word embedding semantic capture (using Wikipedia as the corpus) against semantic relations extracted

from WordNet, as is shown in the results, those vectors achieve good results on semantic capture (94.9%) by themselves. The evaluated semantic relations were *hypernym*, *hyponym*, *holonym* and *meronym*, those relations are one of the most basic semantic relations, which are commonly used on knowledge base construction.

Based on state-of-art classifier model for semantic relation detection, we use Convolutional Neural Network to learn semantic regularities in word embedding space. Based on the results it can be concluded that local regions distribution pattern in word embedding space, where linear semantic regularities are shared between words, can be learned by a non-linear classifier such as CNN. These results bring an interesting case of study, analyze if those distributional patterns can be shared between domains.

7   Future Work

The semantic relation detection using word embedding can be used to infer semantic relations among concepts, and that is highly valuable in areas such as ontology learning and knowledge extraction when dealing with unstructured text data.

One of our interested research domain is ontology learning from unstructured text data in a specialized domain; the result obtained show that might be use semantic relation detection by only word embedding in one of the steps of automatically semantic relations extraction. The results might improve by integrating a decision module that assesses the output by taking advantage of semantic transitivity between relations, consequently, in the semantic relation detection process in knowledge base construction, false positives might be reduced and obtain more reliable knowledge base automatically constructed.

The proposed methodology also can be extended to learn other relations such as *synonym* and *antonym* available in WordNet, also can be interesting research opportunity identify more complex relations using ontologies as a knowledge base to detect relations among entities to discover new knowledge, this can be useful to ontology population and enrichment process.

References

1. Bengio, Y., et al.: A neural probabilistic language model. J. Mach. Learn. Res. **3**, 1137–1155 (2003). doi:10.1162/153244303322533223. ISSN: 15324435. arXiv:1301.3781v3
2. Caraballo, S.A.: Automatic construction of a hypernym-labeled noun hierarchy from text. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, ACL 1999, pp. 120–126. Association for Computational Linguistics, Stroudsburg (1999). ISBN: 1558606093. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.946

3. Chen, Y.-N., et al.: Learning semantic hierarchy with distributed representations for unsupervised spoken language understanding, September 2015

4. Colace, F., et al.: Terminological ontology learning and population using latent Dirichlet allocation. J. Vis. Lang. Comput. **25**(6), 818–826 (2014). doi:10.1016/j.jvlc.2014.11.001. ISSN: 1045926X. http://www.sciencedirect.com/science/article/pii/S1045926X1400127X

5. Fan, M., et al.: Probabilistic belief embedding for large-scale knowledge population. Cogn. Comput. **8**(6), 1–16 (2016). doi:10.1007/s12559-016-9425-5. ISSN: 18669964. arXiv:1505.02433v4

6. Fu, R., et al.: Learning semantic hierarchies: a continuous vector space approach. IEEE Trans. Audio Speech Lang. Process. **23**(3), 461–471 (2015). doi:10.1109/TASLP.2014.2377580. ISSN: 15587916

7. Fu, R., et al.: Learning semantic hierarchies via word embeddings. In: ACL, pp. 1199–1209 (2014)

8. Hendrickx, I., et al.: SemEval-2010 task 8: multi-way classification of semantic relations between pairs of nominals. In: Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions, DEW 2009, pp. 94–99. Association for Computational Linguistics, Stroudsburg (2009). ISBN: 978-1-932432-31-2. http://dl.acm.org/citation.cfm?id=1621969.1621986

9. Hyland, S.L., Karaletsos, T., Rätsch, G.: A generative model of words and relationships from multiple sources. In: Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016), p. 8 (2016). arXiv:1510.00259

10. Kingma, D., Ba, J.: Adam: a method for stochastic optimization, pp. 1–15. arXiv:1412.6980 [cs.LG] (2014)

11. Komninos, A.: Dependency based embeddings for sentence classification tasks. In: Naacl 2016, pp. 1490–1500 (2016)

12. Mikolov, T., Yih, W.-T., Zweig, G.: Linguistic regularities in continuous space word representations. In: Proceedings of NAACL-HLT, pp. 746–751, June 2013. https://www.aclweb.org/anthology/N/N13/N13-1090.pdf

13. Mikolov, T., et al.: Efficient Estimation of Word Representations in Vector Space (2013). http://arxiv.org/abs/1301.3781

14. Qin, P., Xu, W., Guo, J.: An empirical convolutional neural network approach for semantic relation classification. Neurocomputing **190**, 1–9 (2016). doi:10.1016/j.neucom.2015.12.091. ISSN: 18728286

15. Rios-Alvarado, A.B., Lopez-Arevalo, I., Sosa-Sosa, V.J.: Learning concept hierarchies from textual resources for ontologies construction. Expert Syst. Appl. **40**(15), 5907–5915 (2013). doi:10.1016/j.eswa.2013.05.005. ISSN: 09574174

16. dos Santos, C.N., Xiang, B., Zhou, B.: Classifying relations by ranking with convolutional neural networks. In: ACL 2015, vol. 3, pp. 626–634 (2015). arXiv:1504.06580v2. http://arxiv.org/pdf/1504.06580.pdf

17. Takase, S., Okazaki, N., Inui, K.: Modeling semantic compositionality of relational patterns. Eng. Appl. Artif. Intell. **50**, 256–264 (2016). doi:10.1016/j.engappai.2016.01.027. ISSN: 09521976

18. Xiong, S., Ji, D.: Exploiting flexible-constrained K-means clustering with word embedding for aspect-phrase grouping. Inf. Sci. **367–368**, 689–699 (2016). doi:10.1016/j.ins.2016.07.002. ISSN: 00200255