Comparing *Ref-Vectors* and word embeddings in a verb semantic similarity task

Andrea Amelio Ravelli[0000−0002−0232−8881], Lorenzo Gregori[0000−0001−9208−2311], and Rossella Varvara[0000−0001−9957−2807]

University of Florence
andreaamelio.ravelli@unifi.it, lorenzo.gregori@unifi.it, rossella.varvara@unifi.it
http://www.unifi.it

Abstract. This work introduces reference vectors (Ref-Vectors), a new kind of word vectors in which the semantics is determined by the property of words to refer to world entities (i.e. objects or events), rather than by contextual information retrieved in a corpus. Ref-Vectors are here compared with state-of-the-art word embeddings in a verb semantic similarity task. The SimVerb-3500 dataset has been used as a benchmark to verify the presence of a statistical correlation between the semantic similarity derived by human judgments and those measured with Ref-Vectors and verb embeddings. Results show that Ref-Vector similarities are closer to human judgments, proving that, within the action domain, these vectors capture verb semantics better than word embeddings.

Keywords: Ref-Vectors · Verb semantics · Word embedding · Semantic representation

1 Introduction

Natural Language Understanding is a key point in Artificial Intelligence and Robotics, especially when it deals with human-machine interaction, such as instruction given to artificial agents or active learning from observation of human actions and activities. In these scenarios, enabling artificial agents to interpret the semantics and, most of all, the pragmatics behind human speech acts is of paramount importance. Moreover, the verb class is crucial for language processing and understanding, given that it is around verbs that the human language builds sentences.

As an example, consider two of the possible interpretation of a sentence such as *John pushes the glass*: (1) apply a continuous and controlled force to move an object from position A to position B; (2) shove an object away from its position. If we are in a kitchen, in front of a table set with cutlery, plates and glasses, no human will interpret the sentence with (2). This naive example shows how much important is contextual understanding of action verbs.

In this contribution we present a novel semantic representation of action verbs through what we will call *reference vectors* (*Ref-vectors*). Ref-vectors encode

the verb possibility to refer to different types of action. The reference values are extracted from the IMAGACT ontology of action, a multilingual and multimodal ontology built on human competence (sec.3.1). We compared the resulting verb embeddings with other popular word embeddings (namely, word2vec, fastText, GloVe) on the task of word similarity, and benchmarked the results against simVerb-3500 human similarity judgments [4].

2 Related Works

Vector Space Models, also known as Distributional Semantic Models or Word Space Models [25, 7, 21], are widely used in NLP to represent the meaning of a word. Each word corresponds to a vector whose dimensions are scores of co-occurrence with other words of the lexicon in a corpus.

Word embeddings (also known as *neural embeddings*) can be considered the evolution of classical vector models. They are still based on the distributional hypothesis (the semantic of each word is strictly related to word occurrences in a corpus), but are built by means of a neural network that learn to predict words in contexts (e.g. [1], [9]). The literature on the topic is huge and many different models with different features and parameters have been developed through the years ([8]), although only few of them exploit non-contextual features. An important example is the Luminoso system by Speer and Lowry-Duda ([24]) where neural embeddings are enriched with vectors based on common sense knowledge extracted from ConceptNet ([23]).

In this work we also use non-contextual information, but without combining it to traditional word embeddings. Our space is not a co-occurrence matrix but rather a *co-referentiality* matrix: it encodes the ability of two or more verbs to refer to the same action concepts, i.e. local equivalence [17]. Co-referentiality matrices have been successfully used to represent typological closeness of multilingual data [20] and for action types induction [5].

Action concepts in IMAGACT are instantiated as videos depicting actions. The combination of linguistic and visual features to perform a more accurate classification of actions has been widely used in recent years [22, 6, 16], with the development of techniques based on the integration of NLP and computer vision. Within this perspective, our work could be fruitfully exploited to build complex models for action understanding grounded on human knowledge.

We evaluated the performance of the different representation models by means of a verb similarity task, using the SimVerb-3500 dataset [4] as benchmark, which has been previously used in similar works on verbs similarity [2].

3 Action vectors for action verbs

3.1 IMAGACT

IMAGACT[1] [11] is a multimodal and multilingual ontology of action that provides a video-based translation and disambiguation framework for action verbs.

[1] http://www.imagact.it

The resource consists in a fine-grained categorization of action concepts, each represented by one or more visual prototypes in the form of recorded videos and 3D animations. IMAGACT currently contains 1,010 scenes which encompass the action concepts most commonly referred to in everyday language usage. Action concepts have been gathered through an information bootstrapping from Italian and English spontaneous spoken corpora, and the occurrences of verbs referring to physical actions have been manually annotated [13]. Metaphorical and phraseological usages have been excluded from the annotation process, in order to collect exclusively occurrences of physical actions.

The database continuously evolves and at present contains 12 fully-mapped languages and 17 which are underway. The insertion of new languages is obtained through competence-based extension (CBE) [12] by mother-tongue speakers, using a method of ostensive definitions inspired by Wittgenstein [26]. The informants are asked to watch each video, to list all the verbs in their language that correctly apply to the depicted action, and to provide a caption describing the event for every listed verb as an example of use.

The visual representations convey action information in a cross-linguistic environment, offering the possibility to model a conceptualization avoiding bias from monolingual-centric approaches.

3.2 Dataset

The IMAGACT database contains 9189 verbs from 12 languages (Table 1), which have been manually assigned by native speakers to 1010 scenes. It is important to notice that the task has been performed on the whole set of 1,010 scenes for each language and the differences between the number of verbs depend on linguistic factors: some examples of verb-rich languages are (a) Polish and Serbian, in which perfective and imperfective forms are lemmatized as different dictionary entries, (b) German, that have particle verb compositionality, (c) Spanish and Portuguese, for which verbs belong to both American and European varieties.

Judgments of applicability of a verb to a video scene rely on the semantic competence of annotators. An evaluation of CBE assignments has been made for Arabic and Greek [15, 14]; results are summarized in Table 2.

3.3 Creating Ref-vectors

From the IMAGACT database, we derived our dataset as a binary matrix $C_{9189\times1010}$ with one row per verb (in each language) and one column per video prototype. Matrix values are the assignments of verbs to videos:

$$C_{i,j} = \begin{cases} 1 & \text{if verb } i \text{ refers to action } j \\ 0 & \text{else} \end{cases}$$

In this way, the matrix C encodes referential properties of verbs.

In order to provide an exploitable vector representation, an approximated matrix C' has been created from C, by using *Singular Value Decomposition* (SVD) for dimensionality reduction.

Language	**Verbs**
Arabic (Syria)	571
Danish	646
English	662
German	990
Greek	638
Hindi	470
Italian	646
Japanese	736
Polish	1,193
Portuguese	805
Serbian	1,096
Spanish	736
TOTAL	**9189**

Table 1. Number of verbs per language

Language	**Precision**	**Recall**
Arabic (Syria)	0.933	0.927
Greek	0.990	0.927

Table 2. Precision and Recall for CBE annotation task measured on 2 languages

SVD is a widely used technique in distributional semantics to reduce the feature space. The application of SVD to our dataset allowed us to remove the matrix sparsity and to obtain a fixed-size feature space, that is independent of the number of action videos.

Finally, the output C' is a dense matrix 9189×300.

4 Evaluation and results

We compared our co-referentiality vectors to state-of-the-art word embeddings in a verb semantic similarity task. For the evaluation, we considered the full set of 220 English verbs that are shared by the IMAGACT and the SimVerb-3500 dataset. We sampled the comparison dataset (Comp-DS) by taking those pairs of verbs for which similarity scores were present in SimVerb-3500, obtaining 624 verb pairs. Data are reported in Table 3.

Verb semantic similarity has been automatically estimated for each verb pair in the Comp-DS by computing the cosine similarity between the related Ref-vectors. Then, the correlation between automatic and human judgments about verb pair similarity has been determined through the Spearman's rank correlation coefficient[2]. The result is a positive correlation of 0.212 (Table 5). This number highlights a low correlation, but it is not informative without a comparison with other semantic vectors.

[2] We applied the Spearman's rank, because data are non-parametric: Shapiro-Wilk normality test reports $W = 0.9578$ and $p = 1.984\text{e}{-12}$.

	SV-3500	**Comp-DS**
Total verbs	827	220
Total pairs	3500	624
Antonyms	111	34
Cohyponyms	190	57
Hyper/Hyponyms	800	185
Synonyms	306	61

Table 3. Numbers of the full SimVerb-3500 dataset and of the sampled comparison dataset.

To this aim, we considered 6 state-of-the-art word embedding, created with 3 algorithms - word2vec[3] [10], fastText[4] [3] and GloVe[5] [19] - trained on two big corpora - English Wikipedia[6] (2017 dump) and English GigaWord[7] (fifth edition) [18]. In our experiments, we used lemmatized word embeddings, instead of token-specific representation, in order to obtain vectors that are comparable with SimVerb-3500's verb pairs.

Algorithm	**Corpus**	**Lemmas**	**Window**	**Dimensions**
Word2Vec	English Wikipedia 2017	296,630	5	300
Word2Vec	Gigaword 5th Ed.	261,794	5	300
FastText Skipgram	English Wikipedia 2017	273,930	5	300
FastText Skipgram	Gigaword 5th Ed.	262,269	5	300
Global Vectors	English Wikipedia 2017	273,930	5	300
Global Vectors	Gigaword 5th Ed.	262,269	5	300

Table 4. Numbers of the lemmatized word embeddings used for comparison.

The previous procedure has been repeated by using these embeddings instead of Ref-vectors: cosine similarity has been measured between each pair of the Comp-DS, by using different embeddings. The Spearman's rank correlation coefficient with the Comp-DS is reported in Table 5. Data show that Ref-vectors are closer to human judgments in estimating verb semantic similarity (0.212), with the exception of word2vec trained on the GigaW corpus that obtained almost the same degree of correlation (0.211). All the other verb embeddings considered report a lower correlation with Comp-DS.

The same analysis has been conducted considering semantic classes. SimVerb-3500, and thus Comp-DS, contains the annotation of the semantic relation between the two verbs: the pairs can be synonyms, hyper-hyponyms, cohyponyms, antonyms or not related. We used this information to measure the correlation of vector similarity with Comp-DS in verb pairs for each semantic relation. Table 6

[3] https://code.google.com/archive/p/word2vec/
[4] https://fasttext.cc
[5] https://nlp.stanford.edu/projects/glove/
[6] https://archive.org/details/enwiki-20170920
[7] https://catalog.ldc.upenn.edu/LDC2011T07

Ref-vectors	word2vec		fastText		GloVe	
	Wiki	GigaW	Wiki	GigaW	Wiki	GigaW
0.212	0.194	0.211	0.186	0.195	0.105	0.133

Table 5. General correlation results between human judgments from SimVerb-3500 and the compared systems.

shows that Ref-vectors have a stronger correlation (0.26 and 0.35) with human judgments with two classes of related pairs, i.e. hyper-hyponyms and synonyms. Their results are rather poor instead with antonyms and for non semantically related pairs. This suggests that Ref-vectors are better at capturing semantic similarity rather than semantic relatedness. Antonyms, indeed, cannot be considered as semantically similar, since they have opposite meanings. Moreover, if we do not consider pairs that are not semantically related, the general results of Ref-vectors outperform those of the other systems (Table 7).

	Ref-vectors	word2vec		fastText		GloVe	
		Wiki	GigaW	Wiki	GigaW	Wiki	GigaW
Antonyms	0.03	-0.15	0.12	-0.05	0.14	0.09	**0.23**
Cohyponyms	0.20	0.21	0.12	**0.25**	0.13	0.03	-0.02
Hyper-Hyponyms	**0.26**	0.24	0.25	0.23	0.24	0.02	0.04
Synonyms	**0.35**	0.18	0.22	0.13	0.23	0.03	0.14
None	-0.10	**0.16**	**0.16**	0.17	0.15	0.14	0.09

Table 6. Correlation results beetwen systems and simVerb-3500 dataset based on the semantic relation of verb pairs.

Ref-vectors	word2vec		fastText		GloVe	
	Wiki	GigaW	Wiki	GigaW	Wiki	GigaW
0.345	0.174	0.194	0.151	0.171	-0.045	0.029

Table 7. General correlation results for semantically related verb pairs (excluding "None" class)

5 Conclusions

In this paper we have introduced a novel model to represent action verbs semantics based on their referential properties, rather than standard corpus co-occurrences. We compared our model to state-of-the-art word embeddings systems in a verb semantic similarity task. We have shown that our referential vectors correlate better to human judgments from the SimVerb-3500 dataset in presence of specific types of semantic relations. These results suggest that different types of embeddings capture different types of relations, like semantic similarity or simple relatedness. They bring new interesting directions into semantic modeling, a field in which research has focused in the last years mainly on

corpus co-occurrences data. We believe that merging referential and corpus data into semantic modeling can improve language processing and understanding, and foster Natural Language Understanding in Artificial Intelligence towards a more contextually informed dimension.

References

1. Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C.: A neural probabilistic language model. Journal of machine learning research **3**(Feb), 1137–1155 (2003)
2. Blundell, B., Sadrzadeh, M., Jezek, E.: Experimental Results on Exploiting Predicate-Argument Structure for Verb Similarity in Distributional Semantics. In: Dobnik, S., Lappin, S. (eds.) Conference on Logic and Machine Learning in Natural Language (LaML 2017). Gothenburg (2017)
3. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics (TACL) **5**(1), 135–146 (Dec 2017)
4. Gerz, D., Vulic, I., Hill, F., Reichart, R., Korhonen, A.: SimVerb-3500: A Large-Scale Evaluation Set of Verb Similarity. EMNLP (2016)
5. Gregori, L., Varvara, R., Ravelli, A.A.: Action type induction from multilingual lexical features. Procesamiento del Lenguaje Natural **63** (2019)
6. Hahn, M., Silva, A., Rehg, J.M.: Action2Vec: A Crossmodal Embedding Approach to Action Learning. arXiv.org p. arXiv:1901.00484 (Jan 2019)
7. Lenci, A.: Distributional models of word meaning. Annual Review of Linguistics **4**(1), 151–171 (2018)
8. Lenci, A.: Distributional models of word meaning. Annual review of Linguistics **4**, 151–171 (2018)
9. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
10. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. ICLR **cs.CL** (2013)
11. Moneglia, M., Brown, S., Frontini, F., Gagliardi, G., Khan, F., Monachini, M., Panunzi, A.: The imagact visual ontology. an extendable multilingual infrastructure for the representation of lexical encoding of action. In: Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). European Language Resources Association (ELRA), Reykjavik, Iceland (may 2014)
12. Moneglia, M., Brown, S., Kar, A., Kumar, A., Ojha, A.K., Mello, H., Niharika, Jha, G.N., Ray, B., Sharma, A.: Mapping Indian Languages onto the IMAGACT Visual Ontology of Action. In: Jha, G.N., Bali, K., L, S., Banerjee, E. (eds.) Proceedings of WILDRE2 - 2nd Workshop on Indian Language Data: Resources and Evaluation at LREC'14. European Language Resources Association (ELRA), Reykjavik, Iceland (2014)
13. Moneglia, M., Frontini, F., Gagliardi, G., Russo, I., Panunzi, A., Monachini, M.: Imagact: deriving an action ontology from spoken corpora. Proceedings of the Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation (isa-8) pp. 42–47 (2012)
14. Mouyiaris, A.: I verbi d'azione del greco nell'ontologia IMAGACT. Master's thesis, University of Florence (forthcoming)

15. Mutlak, M.: I verbi di azione dell'arabo standard nell'ontologia dell'azione IMAGACT. Ph.D. thesis, University of Florence (2019)
16. Naha, S., Wang, Y.: Beyond verbs: Understanding actions in videos with text. In: Pattern Recognition (ICPR), 2016 23rd International Conference on. pp. 1833–1838. IEEE (2016)
17. Panunzi, A., Moneglia, M., Gregori, L.: Action identification and local equivalence of action verbs: the annotation framework of the imagact ontology. In: Pustejovsky, J., van der Sluis, I. (eds.) Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Paris, France (2018)
18. Parker, R., Graff, D., Kong, J., Chen, K., Maeda, K.: English Gigaword Fifth Edition. Linguistic Data Consortium, LDC2011T07 **12** (2011)
19. Pennington, J., Socher, R., Manning, C.D.: GloVe: Global vectors for word representation. In: Conference on Empirical Methods in Natural Language Processing. pp. 1532–1543. Stanford University, Palo Alto, United States (2014)
20. Ryzhova, D., Kyuseva, M., Paperno, D.: Typology of adjectives benchmark for compositional distributional models. In: Proceedings of the 10th Language Resources and Evaluation Conference. pp. 1253–1257 (2016)
21. Sahlgren, M.: The Word-Space Model. Ph.D. thesis, Stockholm University (2006)
22. Silberer, C., Ferrari, V., Lapata, M.: Models of semantic representation with visual attributes. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 572–582. Association for Computational Linguistics (2013), http://aclweb.org/anthology/P13-1056
23. Speer, R., Chin, J., Havasi, C.: Conceptnet 5.5: An open multilingual graph of general knowledge. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
24. Speer, R., Lowry-Duda, J.: Conceptnet at semeval-2017 task 2: Extending word embeddings with multilingual relational knowledge. arXiv preprint arXiv:1704.03560 (2017)
25. Turney, P.D., Pantel, P.: From frequency to meaning: Vector space models of semantics. Journal of artificial intelligence research **37**, 141–188 (2010)
26. Wittgenstein, L.: Philosophische Untersuchungen. Suhrkamp Verlag (1953)