

Exploring the Suitability of Semantic Spaces as Word Association Models for the Extraction of Semantic Relationships

Epaminondas Kapetanios¹[0000-0002-0617-2183], Vijayan
Sugumaran²[0000-0003-2557-3182], and Anastasia
Angelopoulou¹[0000-0003-1453-492X]

¹ University of Westminster, Regent Street 309, W1B 2HW, London, UK
Regent Street 309, W1B 2HW

{kapetae, agelopa}@westminster.ac.uk

<https://www.westminster.ac.uk/research/groups-and-centres/cognitive-computing-research-group>

² Oakland University, USA
sugumara@oakland.edu

Abstract. Given the recent advances and progress in Natural Language Processing (NLP), extraction of semantic relationships has been at the top of the research agenda in the last few years. This work has been mainly motivated by the fact that building knowledge graphs (KG) and/or bases (KB), as a key ingredient of intelligent applications, is a never-ending challenge, since new knowledge needs to be harvested while old knowledge needs to be revised. Currently, approaches towards relation extraction from text are dominated by neural models practicing some sort of distant (weak) supervision in machine learning from large corpora, with or without consulting external knowledge sources. In this paper, we empirically study and explore the potential of a novel idea of using classical semantic spaces and models, e.g., Word Embedding, generated for extracting word association, in conjunction with relation extraction approaches. The goal is to use these word association models to reinforce current relation extraction approaches. We believe that this is a first attempt of this kind and the results of the study should shed some light on the extent to which these word association models can be used as well as the most promising types of relationships to be considered for extraction.

Keywords: Semantic Relationships · Text Analytics · Natural Language Processing · Distributional Semantics · Word Embedding · Knowledge Graphs · Machine Learning · Algorithms · Artificial Intelligence

1 Introduction

Relationship extraction (RE) is the task of extracting semantic relationships from text. Extracted relationships usually occur between two or more entities of a certain type (e.g. Person, Organisation, Location) and fall into a number

of semantic categories (e.g. married to, employed by, lives in). The importance of the outstanding challenges in extracting semantic relationships from text has been exacerbated by the extraction of knowledge bases and/or graphs as the key ingredient of many AI applications (e.g., word meaning disambiguation algorithms, speech recognition, spell checkers). Moreover, work on RE tasks is motivated by the fact that building such knowledge graphs (KG) and/or bases (KB) is a never-ending challenge because, as the world changes, new knowledge needs to be harvested while old knowledge needs to be revised.

Most work in RE, however, has been witnessed by recent activities showcasing progress in natural language processing, such as SemEval series of competitions for relation extractors. In most of these tasks and approaches, RE is performed by forms of prediction of a relationship by looking either at a short span of text within a single sentence containing a single entity pair mention, or spanning over more than one sentence. In any case, the state-of-the-art in RE builds on neural models using distant (a.k.a. weak) supervision on large-scale corpora for training [30]. Despite the various metrics being used, which makes it difficult to compare systems directly, the range of recall has increased over the years as systems improve, with earlier systems having very low precision at 30% recall. The main metrics used are either precision at N results or plots of the precision-recall.

Since RE tasks and approaches are very similar with tasks known as Knowledge Base or Graph Embedding (KBE), which are concerned with representing KB entities and relations in a vector space for predicting missing links in the graph, attempts to show that combining predictions from RE and KBE models was beneficial for RE [31]. A considerable degree of similarity of RE tasks with classical approaches for the extraction of word associations and meaning, e.g., LSA, LDA, Word2Vec, has been considered as well, since word co-occurrence is the king in these approaches as well.

To the best of our knowledge, however, such latent semantic spaces (LSS) have not been considered as an additional knowledge resource or as a knowledge base to further refine or inform the RE tasks. Combining these two approaches, RE and LSS, could potentially strengthen or weaken the possibility of extracting a semantic relationship, which could be of a more general purpose and not being domain specific.

Having said that, our intention and hope, in this paper, has been to shed some light to the following main question: *Is there a potential correlation between the extractors of semantic relations and the latent semantic spaces for topic modelling and word associations?* Answering this question could help us in further pursuing ways to combine these two approaches in order to use RE for knowledge graph/base refinement and updates.

The paper has been organized as follows. Section 2 provides background and related work, specifically about the knowledge base enhanced relation extraction. Section 3 discusses the methodological and experimental approach. Section 4 presents the preliminary results and concluding remarks.

Contribution: The paper explores for the first time the idea and the possibility of using latent semantic space models to be combined with current approaches for the extraction of semantic relationships in NLP, which are overwhelmingly dominated by neural processing. Using such semantic spaces for word associations, e.g., Word2Vec, could provide good predictions for the kind of semantic relationship we may be looking for.

2 Background and related work

The distinction between RE approaches and those related with the extraction of word associations is not a new one. Generally speaking, since the early 1990s, the line of research around statistical analysis in natural language processing has been split into three main directions: 1) extraction of collocations, which was initiated by Church and Smadja [12], [13], [14] and continued by Evert and Krenn [18], Seretan [19] and Evert [20], with main applications in translation and language teaching, 2) extraction of word associations and computation of semantic similarities [1], [2], [3], [4], and 3) (semi-)automatic extraction of particular linguistic relations (or thesaurus relations), which are also known as automatic construction of a thesaurus.

The third direction of research is attributed to the (semi-)automatic extraction of particular linguistic relations (or thesaurus relations), e.g., [21], which are also known as automatic construction of a thesaurus. This line of development has been distinguished from the other two lines of research in that it introduces a different methodology based on second order statistics, differentiating between syntactic and paradigmatic relations [22], context comparisons [23]. Besides, this line of development attempts to give the term word association a more precise definition, which can be used to denote various kinds of linguistic relations, often synonyms, sometimes plain word association (play, soccer) and sometimes other linguistic relations like derivation and hyperonymy, antonyms, qualitative direction of adjectives (negative vs. positive). Word sense distinction, contrary to word sense disambiguation, belongs to this area as well, since it describes just another kind of specific relation between words.

All these approaches, however, do rely on the distributional hypothesis, a mathematically motivated line of influence on today's computation of relations between words as firstly established by Zelig Harris [11]. Another common feature has been that the main goal was to provide information on the general combinatorial possibilities of an entry word. Various types of combinatorial preferences are listed, such as, whether there are any combinatorial preferences of verbs for nouns (e.g. [to adopt, enact, apply] a regulation) or what the possible adverbial combinations (i.e. modifications) of a verb are (e.g. to regret [deeply, very much]). There is also a distinction between grammatical and lexical collocations with the latter relying on part-of-speech patterns, such as verb-(preposition)-noun, adjective-noun or noun-noun, for permissible collocations in a natural language. For instance, compose music and launch a missile are permissible, while compose a missile is at least awkward.

Generally speaking, associations have been distinguished as *association by similarity*, *contrast* and *contiguity*. Association by similarity is based on the fact that the associated phenomena have some kind of common features. Association by contrast has its origin in what is explained by the presence in phenomena of opposite features. For example, the phenomenon of antonyms: grief - joy, happiness - unhappiness, and so on. Association by contiguity comes into existence when events are situated close together in time or space. Along with them, more complex semantic associations are distinguished. These are, in particular, the association reflecting generic and cause-and-effect relationship between the objects of the world, e.g., a flower - a rose, a disease - death, and so on.

It is also well acknowledged that association is one of the basic mechanisms of memory. In a sense, these mechanisms can be called natural classifiers of the conceptual content of the vocabulary of the language. Ideas and concepts, which are available to the memory are related. This relationship is based on human past experience and, more or less, accurately reproduces an objectively existing relationship between the phenomena of the real world. Under certain conditions, a revival of one idea or concept is accompanied by a revival of other ideas correlated with it. This phenomenon is called the association (a term proposed in the XVIII century by Locke).

In this context, the usage of the term word association indicates a broader meaning. In their examples of automatically computed, strongly associated word pairs, there is a mentioning of semantic relations such as meronymy, hyperonymy and so forth. Smadja, however, mentions them as examples of where Church's algorithm computed just pairs of words that frequently appear together [15]. Lin [16] even considers doctors and hospitals as unrelated and thus wrongly computed as significant by Church and Hanks, although they stand in a meronymy relation. Nonetheless, other contemporaries, e.g., Dunning [17], improved the mathematical foundation of this research field by introducing the log-likelihood measure. Dunning was the first to coin the term statistical text analysis.

Despite all these differences, the commonalities between the two worlds, i.e., current RE tasks and approaches, e.g., HRERE [32], and word association extraction models have not been explored further. In HRERE, however, there is only attempt to combine knowledge bases with neural networks targeting relation extraction. A much tighter integration of RE and KBE models is needed with the purpose of using them not only for prediction, but also train them together, thus mutually reinforcing one another. Several other methods have also been proposed [33], [34], [35], [36] to use information from KBs to facilitate relation extraction. These vary from considering other relations in the sentential context while predicting the target relation, to utilising additional side information from KBs for improved RE.

In our paper, we take a different approach in that we consider and investigate the possibility, for the first time, to use latent semantic spaces, such as LSA and Word2Vec (pre-trained and no training), for the extraction of word associations as a knowledge base to inform or reinforce relation extraction approaches. Despite the fact that current RE extractors are reporting various metrics, making

it difficult to compare systems directly, the main metrics to be used by our approach will be precision at N results or plots of the precision-recall as well and in line with most of the RE approaches.

3 Methodology and experimentation

In order to explore the merits of the idea and proposal to use well established semantic spaces for the extraction of word associations for reinforcement of current extractors of semantic relationships, the experimental setup was very much aligned with the SemEval-2010 Task 8, Multi-Way Classification of Semantic Relations between Pairs of Nominals. The task has been, given a sentence and two tagged nominals, to predict the relation between those nominals and the direction of the relation. The dataset contains nine general semantic relations together with a tenth OTHER relation.

For instance, given the sentence:

There were apples, pears and oranges in the bowl. the semantic relationship

(content-container, pears, bowl) should be derived.

In fact, we used OpenNRE as an open-source and extensible toolkit that provides a unified framework to implement relation extraction models. Subsequently, we applied it to the Google News data set and corpus, in order to extract relations from this corpus. The reason for that is merely the fact that the same corpus has also been used for the extraction of the published pre-trained vectors (e.g., Word2Vec pre-trained) being used for extraction of word associations. This model contains 300-dimensional vectors for 3 million words and phrases. The archive is available at <https://code.google.com/archive/p/word2vec/>.

Apart from the pre-trained Word2Vec model, we also considered three experimental implementations of LSA, LDA and Word2Vec (no training) from scratch in Python by using Gensim and NumPy, libraries for the first two semantic spaces and the third one, respectively. The intention has been to bring these semantic spaces generated by these classical approaches for the extraction of word associations and meaning into consideration as well.

In order to explore the potential of an interesting correlation between RE tasks and knowledge provided by such semantic spaces, we inputted all the extracted relations as a pair of entities into the four (4) semantic spaces. This would allow us to see whether their association indicated by the RE approach could be reproduced within these semantic spaces. To achieve an objective picture of the relations distribution within the semantic spaces being extracted, we decided to run the experiment with 100 randomly chosen terms underpinned by nouns.

The results of this empirical study are presented based on the classical precision-recall tandem of metrics as follows:

Relation Count (RC): *Number of relations identified within the semantic space. This is an aggregate of all relations being returned/reproduced within the 10 closest terms and for all input terms.*

Semantic Space Relation Inclusion (SSRIC): *TR / N , N =number of extracted relations from OpeNRE, TR number of input terms, which return at least one relation within the 10 closest.*

Relation Precision (R-Prec): *Number of correct relations, meaning that these relations have been identified as associations. In other words: $R\text{-Prec} = \text{Number of all retrieved relations} / \text{Number of all correctly identified relations}$.*

Relation Recall (R-Rec): *Number of retrieved relations, meaning that these relations have been identified as associations with the input term. In other words: $R\text{-Rec} = \text{Number of all retrieved relations} / \text{Number of all possible relations}$.*

4 Experimental results and discussion

The following table 1 summarizes the first results from this empirical study.

Table 1. Precision-recall for reproducing extracted relationships within the semantic similarity spaces for word associations

	RC	SSRIC	R-Prec	R-Rec
LSA	210	0.45	0.37	0.21
LDA	245	0.55	0.45	0.24
Word2Vec	235	0.65	0.48	2.4
Word2Vec (pre-trained)	270	0.65	0.49	2.7

The following tables shed more light on precision-recall based reproduction of certain categories of extracted relationships within the extracted semantic spaces for word associations.

Cause-Effect Relations	N / Total	R-Prec	R-Rec
LSA	10 / 210	0.24	0.11
LDA	9 / 245	0.35	0.12
Word2Vec	7 / 235	0.28	0.115
Word2Vec (pre-trained)	13 / 270	0.29	0.17

Component-Whole Relations	N / Total	R-Prec	R-Rec
LSA	22 / 210	0.26	0.18
LDA	25 / 245	0.36	0.19
Word2Vec	26 / 235	0.29	0.195
Word2Vec (pre-trained)	28 / 270	0.33	0.198

Content-Container Relations	N / Total	R-Prec	R-Rec
LSA	27 / 210	0.28	0.28
LDA	24 / 245	0.26	0.39
Word2Vec	23 / 235	0.39	0.195
Word2Vec (pre-trained)	26 / 270	0.43	0.198

Entity-Destination	N / Total	R-Prec	R-Rec
LSA	17 / 210	0.18	0.18
LDA	14 / 245	0.16	0.19
Word2Vec	13 / 235	0.29	0.195
Word2Vec (pre-trained)	16 / 270	0.23	0.198

Entity-Origin	N / Total	R-Prec	R-Rec
LSA	25 / 210	0.36	0.28
LDA	28 / 245	0.36	0.29
Word2Vec	29 / 235	0.39	0.24
Word2Vec (pre-trained)	28 / 270	0.4	0.26

Message-Topic	N / Total	R-Prec	R-Rec
LSA	8 / 210	0.28	0.38
LDA	25 / 245	0.46	0.49
Word2Vec	9 / 235	0.32	0.24
Word2Vec (pre-trained)	8 / 270	0.33	0.25

Member-Collection	N / Total	R-Prec	R-Rec
LSA	48 / 210	0.48	0.35
LDA	45 / 245	0.49	0.39
Word2Vec	49 / 235	0.42	0.34
Word2Vec (pre-trained)	58 / 270	0.43	0.35

Instrument-Agency	N / Total	R-Prec	R-Rec
LSA	5 / 210	0.21	0.28
LDA	10 / 245	0.26	0.19
Word2Vec	6 / 235	0.22	0.14
Word2Vec (pre-trained)	8 / 270	0.13	0.15

Product-Producer	N / Total	R-Prec	R-Rec
LSA	12 / 210	0.31	0.35
LDA	15 / 245	0.36	0.39
Word2Vec	19 / 235	0.32	0.29
Word2Vec (pre-trained)	18 / 270	0.3	0.28

Other	N / Total	R-Prec	R-Rec
LSA	36 / 210	0.16	0.18
LDA	50 / 245	0.16	0.19
Word2Vec	54 / 235	0.12	0.14
Word2Vec (pre-trained)	67 / 270	0.13	0.15

4.1 Discussion of results

Interesting patterns emerging from this first analysis have been summarized as follows.

- Any extracted relations, used as input into the four semantic space models for word associations, do not appear as being among the 10 closest terms for more than 65% of the input pairs (best case with the Word2Vec pre-trained model).
- Following the break down of counts of different extracted relations identified, the highest scores have been encountered for the relation type *member-collection*, which is typical for associations of words where hyponymy or hyperonymy (lexical semantics) is hidden. This, in turn, may indicate that these types of relationships are likely to be reproduced within the considered semantic spaces for word associations.
- The lowest scores being encountered for the types of relations *Instrument-Agency*, *product-producer* and *Entity-Destination* may denote that these types of relationships can hardly be reproduced within the considered semantic spaces for word associations.
- Overall, one may also set up a further hypothesis that many of the detected relations by word association semantic spaces and models, remain undetected by the RE extractors, if they are not combined.

5 Conclusion

In this paper, we took a novel idea to combine knowledge extracted in form of word associations with some classical approaches, e.g., LSA, Word2Vec, with current approaches for semantic relationship extraction (RE), to the test. In particular, we conducted an experiment and empirical study to test the hypothesis whether the extracted relations, in the form of pairs of entities or concepts among which a relation has been extracted, returned by current RE approaches can be reproduced within word association semantic spaces. Studying this overlap could shed some light on the potential of a correlation between the two worlds, which have never been used in combination so far. Our aspiration is to inform our RE approach, currently under development, by such word association models and semantic spaces. We believe that certain types of relations, at a more generic level, can be easily detected by combing the two worlds.

References

1. Levy, O., Goldberg, Y.: Neural word embedding as implicit matrix factorization. In Advances in neural information processing systems, pp. 2177–2185, (2014)

2. Artetxe, M., Labaka, G., Lopez-Gazpio, I., Agirre, E.: Uncovering divergent linguistic information in word embeddings with lessons for intrinsic and extrinsic evaluation. arXiv preprint arXiv:1809.02094, (2018)
3. Arora, S., Li, Y., Liang, Y., Ma, T., Risteski, A.: A latent variable model approach to PMI based word embeddings. *Transactions of the Association for Computational Linguistics*, 4, pp. 385–399, (2016)
4. Gittens, A., Achlioptas, D., Mahoney, M. W.: Skip-gram-zipf+ uniform=vector additivity. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 69-76, (2017)
5. Manning, C. D., Schtze, H.: *Foundations of Statistical Natural Language Processing*; MIT Press, (1999)
6. Firth, J. R.: *Papers in Linguistics 1934–1951*; Oxford University Press: London, U.K. (1957)
7. Benson, M., Benson, E., Ilson, R.: *The BBI combinatory dictionary of English: A guide to word combinations*; John Benjamins: Amsterdam, (1986)
8. Benson, M.: The structure of the collocational dictionary. *International Journal of Lexicography*, 2(1), pp. 1-14, (1989)
9. Benson, M.: Collocations and general-purpose dictionaries. *International Journal of Lexicography*, 3(1), pp. 23-35, (1990)
10. Kent, G., Rosanoff, A.J.: A study of association in insanity. *American Journal of Insanity*, 67, pp. 317–390, (1910)
11. Zeelig, H. S.: *Mathematical Structures of Language*, Wiley, New York (1968)
12. Choueka, Y., Klein, S. T., Neuwitz, E.: Automatic retrieval of frequent idiomatic and collocational expressions in a large corpus. *Journal of the Association for Literary and Linguistic Computing*, pp. 34-38, (1983)
13. Church, K. W., Gale, W. A., Hanks, P., Hindle, D.: Using statistics in lexical analysis. In Zernik, U. (eds.) *Lexical Acquisition: Exploiting On-Line Resources to Build up a Lexicon*, pp. 115-164, (1991)
14. Smadja, F. Macro-coding the lexicon with co-occurrence knowledge. In: Zernik, U. (eds.) *Proceedings of the First International Lexical Acquisition Workshop*, (1989)
15. Smadja, F. A.: Retrieving collocations from text: Xtract. *Computational Linguistics*, 19(1), pp. 143-177, (1993)
16. Lin, D.: Extracting collocations from text corpora. In: *Proceedings of the 1st Workshop on Computational Terminology*, Montreal, Quebec, Canada (1998) pp. 57-63.
17. Dunning, T.: Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), pp. 61-74, (1993)
18. Evert, S., Krenn, B.: Methods for the qualitative evaluation of lexical association measures. In: *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pp. 188-195, Toulouse, France (2001)
19. Seretan, M.-V.: *Syntactic and Semantic Oriented Corpus Investigation for Collocation Extraction, Translation and Generation*. Ph.D. thesis, Language Technology Laboratory, Department of Linguistics, Faculty of Arts, University of Geneva (2003)
20. Evert, S.: *The Statistics of Word Co-occurrences: Word Pairs and Collocations*, Ph.D. thesis, University of Stuttgart (2005)
21. Ruge, G.: Automatic detection of thesaurus relations for information retrieval applications. In: Freksa, C., Jantzen, M., Valk, R., (eds.) *Foundations of Computer Science: Potential - Theory Cognition*, pp. 499-506, Springer Verlag, Heidelberg
22. Rapp, R.: The computation of word associations. In: *Proceedings of COLING-02*, Taipei, Taiwan (2002)

23. Biemann, C., Bordag, S., Heyer, G., Wolff, C. Language-independent methods for compiling monolingual lexical data. In: Proceedings of CICLing, Springer Verlag, pp. 215–228, (2004)
24. Pantel, P., Lin, D.: Word-for-word glossing with contextually similar words. In Proc. of the 1st Annual Meeting of the North American Chapter of Association for Computational Linguistics, pp. 78-85, Seattle, USA (2000)
25. Schtze, H.: Automatic word sense discrimination. *Computational Linguistics*, 24, pp. 97-124, (1998)
26. Grefenstette, G.: *Explorations in Automatic Thesaurus Discovery*, Kluwer Academic Press, Boston (1994)
27. Matsumura, N., Ohsawa, Y., Ishizuka, M.: PAI: Automatic indexing for extracting asserted keywords from a document. *New Generation Computing*, 21(1), pp. 37-47, (2003)
28. Salton, G., Singhal, A., Mitra, M., Buckley, C.: Automatic text structuring and summarization. *Information Processing and Management*, 33(2), pp. 193-207, (1997)
29. Witschel, F.: Terminology extraction and automatic indexing - comparison and qualitative evaluation of methods. In: Proc. of Terminology and Knowledge Engineering, (2005)
30. Mintz, M., Bills, S., Snow, R., Juraf-sky, D.: Distant supervision for relation extraction without labeled data. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2, pages 1003-1011. Association for Computational Linguistics, (2009)
31. Weston, J., Bordes, A., Yakhnenko, O., Usunier, N.: Connecting language and knowledge bases with embedding models for relation extraction. arXiv preprint arXiv:1307.7973, (2013)
32. XU, P. Barbosa, D.: Connecting Language and Knowledge with Heterogeneous Representations for Neural Relation Extraction. <http://arxiv.org/abs/1903.10126>, (2019)
33. Sorokin, D., Gurevych, I.: Context-aware representations for knowledge base relation extraction. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1784-1789, (2017)
34. Vashishth, S., Joshi, R., Prayaga, S. S., Bhattacharyya, C., Talukdar, P.: Reside: Improving distantly-supervised neural relation extraction using side information arXivpreprint arXiv:1812.04361. (2018)
35. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Advances in neural information processing systems, pages 2787-2795. (2013)
36. Han, X., Liu, Z., Sun, M.: Neural knowledge acquisition via mutual attention between knowledge graph and text. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)