

# Estimating Text Similarity based on Semantic Concept Embeddings

Tim vor der Brück

School of Computer Science and Information Technology  
Lucerne University of Applied Sciences and Arts (HSLU)  
Rotkreuz, Switzerland  
tim.vorderbrueck@ffhs.ch

Marc Pouly

School of Computer Science and Information Technology  
Lucerne University of Applied Sciences and Arts (HSLU)  
Rotkreuz, Switzerland  
marc.pouly@hslu.ch

**Abstract**—Due to their ease of use and high accuracy, Word2Vec (W2V) word embeddings enjoy great success in the semantic representation of words, sentences, and whole documents as well as for semantic similarity estimation. However, they have the shortcoming that they are directly extracted from a surface representation, which does not adequately represent human thought processes and also performs poorly for highly ambiguous words. Therefore, we propose Semantic Concept Embeddings (CE) based on the MultiNet Semantic Network (SN) formalism, which addresses both shortcomings. The evaluation on a marketing target group distribution task showed that the accuracy of predicted target groups can be increased by combining traditional word embeddings with semantic CE.

**Index Terms**—Concepts, MultiNet, concept embeddings, semantic similarity estimation.

## I. INTRODUCTION

Word2Vec (W2V) word embeddings became a popular method for creating semantic representations of words, sentences, and whole documents as well as their semantic similarity estimation. Their key success factors are ease of use, scalability, and performance. However, there are also important downsides. Consider, for example, a word just in front of an embedded sub-clause. Potentially, the words just succeeding this sub-clause might constitute a more characteristic word context than the words actually occurring inside this sub-clause despite being farther apart. So, in some situations, considering only the word context based on the surface structure may not be the optimal choice. Hence, we hereby introduce semantic Concept Embeddings (CEs) that consider the neighborhood in a semantic network (SN). We expect a meaning-oriented structure such as an SN to provide a much more adequate description of neighborhood since it is oriented on human thought processes. The SNs are automatically constructed from arbitrary surface texts by the Wocadi parser [1], which combines manual word function-oriented rules with statistical disambiguation methods. In a pre-study, Wocadi obtained superior results than the only state-of-the-art semantic role labeling parser supporting German out-of-the-box, which is AMR-Eager-Multilingual [2] (AMR stands for Abstract Meaning Representation), a Deep Learning based approach. The SNs as generated by Wocadi are much more comprehensive than, for example, WordNet [3] or GermaNet [4], since they can represent the semantics of arbitrary texts and not only ontologies.

The CEs obtained from these SNs are then employed in the task of assigning participants of an online contest to certain

marketing target groups, called youth milieus, by analyzing short text snippets provided by these participants.

Finally, we applied our estimate to the scenario of distributing participants of an online contest into several target groups called youth milieus by exploiting short text snippets they were asked to provide. The evaluation revealed that our novel estimators performed superior to several baseline methods for this scenario.

The remainder of the paper is organized as follows. In the next section, we look into several state-of-the-art methods for estimating semantic similarity. Sect. III introduces the MultiNet semantic network formalism. In Sect. IV, we describe in detail how these networks can be used for estimating semantic similarity between two texts. Our application scenario for our proposed semantic similarity estimates is given in Sect. V. Sect. VI describes the conducted evaluation, in which we compare our approach with several baseline methods. The results of the evaluation are discussed in Sect. VI. Finally, this paper concludes with Sect. VII, which summarizes the obtained results.

## II. RELATED WORK

The main application area of knowledge graph embeddings is link prediction [5]. They are very rarely used to estimate the semantic similarity of texts, confer as an example for the latter the approach of Goikoetxea et al. [6]. They generate random walks on WordNet to extract sequences of words, where the lexicalization is randomly chosen for the associated synset nodes. These sequences are then fed into the ordinary W2V to create (ontology) embedding vectors. They evaluated several possibilities to combine such vectors with ordinary word embeddings obtained from large corpora like averaging or concatenating them. Another embeddings extraction method for WordNet synsets is proposed by Kutuzov et al. [7]. They obtained superior results to random walks by obtaining the embedding vectors as a result of an optimization problem that ensures local (graph neighbor) and global (user annotations) consistency. Note that WordNet is much smaller than Wikipedia, which we use as a basis for our approach. Furthermore, MultiNet concepts have some advantages over MultiNet synsets, since in some cases word senses cannot be fully captured by synsets.

An approach to obtain CEs from ordinary surface texts is proposed by Mencia et al. [8]. In particular, they generate CE by counting co-occurrences of *concepts* akin to obtaining

GloVe [9] word embeddings. A similar approach was introduced by Beam et al. [10] that directly uses GloVe to obtain the semantic concept vectors. However, both approaches do not employ a proper Word Sense Disambiguation to obtain their concept representation. Instead, they restrict their embeddings vocabulary to certain medical terms, which are usually specific enough to represent only unique word senses.

After their extraction, word or concept embeddings can be used to estimate document similarity as follows:

- 1) The embeddings (often weighted by the tf-idf coefficients of the associated words [11]) are looked up in a hashtable for all the words in the two documents to compare. These embeddings are determined beforehand on a very large corpus typically using either the skip-gram or the continuous bag of words variant of the W2V model [12]. The skip-gram method aims to predict the textual surroundings of a given word by means of an artificial neural network. The influential weights of the one-hot-encoded input word to the nodes of the hidden layer constitute the embedding vector. For the so-called *continuous bag of words* method, it is just the opposite, i.e., the center word is predicted by the words in its surroundings.
- 2) The centroid over all word embeddings belonging to the same document is calculated to obtain its vector representation.

Alternatives to W2V are GloVe [9], which is based on aggregated global word co-occurrence statistics and the Explicit Semantic Analysis (or shortly ESA) [13], in which each word is represented by the column vector in the tf-idf matrix over Wikipedia.

The idea of W2V can be transferred to the level of sentences as well. In particular, the so-called Skip Thought Vector (STV) model [14] derives a vector representation of the current sentence by predicting the surrounding sentences.

If vector representations of the two documents to compare were successfully established, a similarity estimate can be obtained by applying the cosine measure to the two vectors. [15] propose an alternative approach for ESA word embeddings that establishes a bipartite graph consisting of the best matching vector components by solving a linear optimization problem. The similarity estimate for the documents is then given by the global optimum of the objective function. However, this method is only useful for sparse vector representations. In the case of dense vectors, [16] suggests applying the Frobenius kernel to the embedding matrices, which contain the embedding vectors for all document components (usually either sentences or words, cf. also [17]). However, crucial limitations are that the Frobenius kernel is only applicable if the number of words (sentences respectively) in the compared documents coincide and that a word from the first document is only compared with its counterpart from the second document. Thus, an optimal matching has to be established already beforehand.

Word embeddings as described so far are represented by a fixed word-vector mapping, which means that these word vectors do not vary depending on the current word context.

This is different for Elmo [18] and Bert [19], [20] embeddings that take this context into account basically realizing a kind of word sense disambiguation.

Before going into more details of our method, we first want to introduce the MultiNet SN formalism.

### III. SNS BASED ON THE MULTINET FORMALISM

In contrast to other popular knowledge graphs or SNs like WordNet [3], OdeNet [21] or Yago [22] that represent ontologies, MultiNet is designed to grasp the entire semantics of natural language. Therefore, MultiNet embeddings can be trained on much larger data sets than the formerly mentioned knowledge graphs, in case of this paper on the entire German Wikipedia. SNs of the MultiNet (Multilayered Extended SNs) formalism [23] allow to homogeneously represent the semantics of single words, phrases, sentences, texts, or text collections.

An SN node represents a concept, while an SN arc expresses a relation between two concepts. A concept *lemma.x.y* is represented by a lemma, and two numbers, where the first (x) denotes the homograph and the second (y) the polyseme. In addition, each node is semantically classified by a *sort* from a hierarchy of 45 sorts organized in a taxonomy.

Furthermore, a node has an inner structure (depending on its sort) containing *layer features* like CARD (cardinality) and REFER (referential determinacy).

The Wocadi parser can construct SNs of the MultiNet formalism for German phrases, sentences, or texts. During this process, SNs and syntactic dependency structures are built.

An important component of this deep syntactico-semantic analysis of natural language is HaGenLex, a semantically based computer lexicon [24]. This lexicon not only lists verb valencies, but also their syntactic and semantic types. Consider, for example, the German verb *essen* ('eat'). Sentences like *Die Birne isst den Apfel.* ('The pear eats the apple.') are rejected because semantic selectional restrictions are violated. Besides this comprehensive lexicon with around 28,000 entries, a shallow lexicon, many name lexicons, and a sophisticated compound analysis is applied to achieve the parser coverage required for natural language processing applications.

Disambiguation is realized by specialized modules which work with symbolic rules and disambiguation statistics derived from annotated corpora. Currently, such modules exist for (intrasentential and intersentential) coreference resolution, the attachment of prepositional phrases, and the interpretation of prepositional phrases.

The following MultiNet relations and functions are relevant to this paper [23]:

- AGT: Definition: In its standard interpretation, the expression (e AGT o) characterizes the relation between an event e or, to be specific, an action e and a conceptual object o which actively causes e (i.e. o is originating/sustaining/giving rise to e). In other words, o is the active object (the agent or carrier of the action).

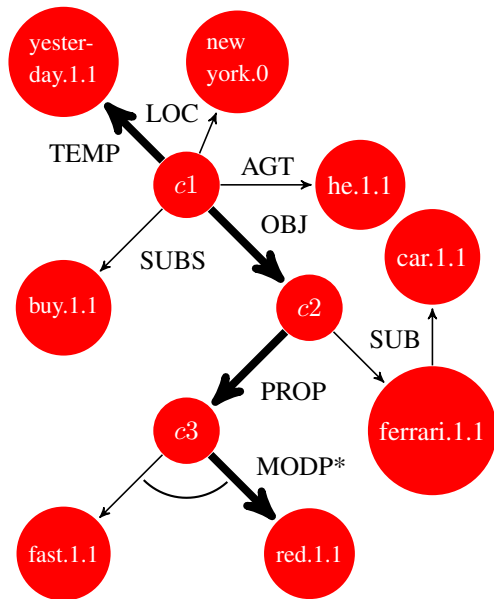


Fig. 1. Example for a random walk (yesterday.1.1 TEMP<sup>-1</sup> OBJ PROP MODP\* red.1.1) in the SN.

- CHEA: The statement (e CHEA a) expresses the connection between an event e (usually a verb) and an abstract concept a (usually a noun) which agree, at least partially, in their meaning. Concepts connected by CHEA correspond to each other in a systematic way. Example: (inform.1.1 CHEA information.1.1)
- CHPA: The statement (p CHPA c) establishes a connection between a property p (usually adjective) and an abstract concept (usually a noun) c which is semantically close to p and whose meaning is systematically related to p. Example: (cold.1.1 CHPA coldness.1.1).
- SUB: The expression (o<sub>1</sub> SUB o<sub>2</sub>) specifies that the individual or generic concept o<sub>1</sub> is subordinate (a hyponym) to the generic concept o<sub>2</sub>, i.e. everything derivable for o<sub>2</sub> is also valid for o<sub>1</sub>.

An example for an SN following the MultiNet formalism is given in Figure 1.

#### IV. USING SEMANTIC CEs FOR ESTIMATING SEMANTIC SIMILARITY

##### A. Procedure

To better understand why a representation as an SN can be beneficial, let us consider the following example sentence: *The car he bought yesterday in New York is a fast red Ferrari.* This sentence contains the embedded relative clause *he bought yesterday in New York*. Actually, the noun phrase *fast red Ferrari* is much more characteristic for the word *car* than the subclause *he bought yesterday in New York* despite being farther away from that word. The representation of this sentence as an SN on MultiNet is shown in Figure 1. The main node representing the entire sentence is c1, which is a specific buying operation expressed by the relation SUBS. The relation LOC specifies the location of the event, in this case *New York*.

TABLE I  
COSINE MEASURE APPLIED TO WORD AND CONCEPT VECTORS.

Wort 1 / Concept 1	Word 2 / Concept 2	Example for Concept 2	Cosine
Lok (locomotive)	Zug (move(ment)/train)		0.428
lok.1.1	zug.1.1	Der Zug nach London fährt ab. (The train to London is departing.)	0.633
lok.1.1	zug.1.2	Zug der Vögel (bird migration (movement))	0.321

The trailing .0 at the concept name denotes the fact that the concept is identified by a proper name. The agent (relation: AGT) of the event is *he.1.1*, which can potentially be resolved by an anaphora resolution and replaced by its antecedent. The Ferrari that is bought is identified by the concept c2, connected to c1 by an object (OBJ) relation. The red Ferrari is also fast, which is expressed by a modifier (MODP\*) combined with a property relation (PROP).

Now let us consider a second example given by the following two sentences: *Pete kauft einen neuen roten Ferrari.* (*Pete buys a new red Ferrari.*) und *Ein neuer roter Ferrari wird von Pete gekauft.* (*A new red Ferrari is bought by Pete.*) Both sentences express the same meaning but the surface structure is different. The representation in passive voice contains the additional word (*wird* (English: *was*)), which is in German a function word that can also express future tense, i.e., this word is ambiguous and can therefore add additional noise to the word vectors. The SNs of both sentences however coincide, which results in more exact embedding vectors.

Let us consider as a third example the word *space*, which can either denote a certain geographical area or the universe. Actually, the word vector of *space* integrates all possible word senses. Thus, the cosine distance between the word vectors of *space* and *planet* should be smaller than for *universe* and *planet*. However, if it is known from the word context that *space* means *universe* in a certain text, then *space* and *planet* should be similarly semantically related as *universe* and *planet*. Using our SN-based approach, we no longer generate vectors for surface words but for word senses instead, which solves this issue. A concrete German example comprises the two German words *Lok* (locomotive) and *Zug* (move(ment) / train) and is given in Table I. It demonstrates that the concept *lok.1.1* (locomotive.1.1) has a higher cosine similarity to the word sense of *Zug* denoting *train* (*zug.1.1*) than to the word sense *zug.1.2* denoting *move(ment)*, and this cosine similarity also exceeds the one between the word vectors of *Zug* and *Lok*.

Our workflow to create semantic CEs is as follows:

- Parse the German Wikipedia by means of Wocadi to create SNs
- Create random walks on these SNs

- Feed the random walks into W2V to create CEs

Note that we obtained already a fully parsed Wikipedia from Sempria GmbH, a spin-off of the Distance University of Hagen. Once the CEs are generated, they can be exploited to estimate semantic text similarity in the following way:

- Apply a Word Sense Disambiguation to both texts to obtain MultiNet concepts.
- Create CEs for both concept representations.
- Determine the two CEs centroids.
- Estimate the semantic similarity between two texts using the cosine measure.

There are several approaches to obtaining embeddings from knowledge graphs and SN like RotatE [25], TransE [26], or TorusE [27]. We decided to train W2V on random walks over the MultiNet-SNs, which has the advantage that hereby word embeddings and CEs are better comparable since both are obtained using the same method, namely W2V. To generate such a random walk, one first picks an initial node from an SN randomly. Afterwards, one randomly chooses a neighboring node and repeats this process using the newly picked node. Additionally, one generates in each step a continuous random number between 0 and 1 and terminates the random walk, if this number assumes a value below a certain (small) threshold. Note that in contrast to a path, edges can appear several times in a certain random walk and one can also move back and forth on the same edge. The random walk representation is then given by an interleaving sequence of nodes and edges. If an edge is followed against the arc direction, we note down the inverse of the relation associated with the arc. Let us consider as an example the SN shown in Figure 1, in which the following possible random walk is already highlighted:

yesterday.1.1 TEMP<sup>-1</sup> c1 OBJ c2 PROP c3 MODP\* red.1.1

Since the inner nodes like c1 and c2 are named arbitrarily without expressing any sort of meaning, we remove such nodes from the random walk. Thus, our modified random walk becomes:

yesterday.1.1 TEMP<sup>-1</sup> OBJ PROP MODP\* red.1.1

In case of symmetric relations  $Rel$  with  $Rel = Rel^{-1}$ , we only consider the non-inverted relations in the random walk. In our example, the random walk does not contain any symmetric relations so it stays unchanged.

To estimate the semantic similarity between two texts using CEs, one first has to map the words appearing in these texts to concepts, which boils down to applying a Word Sense Disambiguation. One possibility to accomplish this task is to parse both texts using Wocadi, which generates for each text an SN and also establishes a mapping from each word to the associated concept. However, note that Wocadi is not freely available but requires a license from Sempria GmbH. Therefore, we use a different approach based on so-called word-concept embeddings. As the name insinuates, word-concept embeddings denote concept vectors, which are based on a surface oriented approach. In particular, we obtain a word-concept vector of some concept  $c$  by averaging the word vector centroids of all sentences, where  $c$  occurs in the associated SN. We then chose the word sense  $c$  of a word  $w$  occurring in a

certain sentence  $s$ , where the word-concept vector of  $c$  is most similar to the centroid of  $s$ .

## V. APPLICATION SCENARIO: MARKET SEGMENTATION

Market segmentation is one of the key tasks of a marketer. Usually, it is accomplished by clustering over behaviors as well as demographic, geographic, and psychographic variables [28]. In this paper, we will describe an alternative approach based on unsupervised natural language processing. In particular, our business partner operates a commercial youth platform for the Swiss market, where registered members get access to third-party offers such as discounts and special events like concerts or castings. Actually, several hundred online contests per year are launched over this platform sponsored by other firms, an increasing number of them require the members to write short free-text snippets, e.g. to elaborate on a perfect holiday at a destination of their choice in case of a contest sponsored by a travel agency. Based on the results of a broad survey, the platform provider's marketers assume five different target groups (called *youth milieus*) being present among the platform members: *progressive postmodern youth* (people primarily interested in culture and arts), *young performers* (people striving for a high salary with a strong affinity to luxury goods), *freestyle action sportsmen*, *hedonists* (rather poorly educated people who enjoy partying and disco music) and *conservative youth* (traditional people with a strong concern for security). A sixth milieu called *special groups* comprises all those who cannot be assigned to one of the upper five milieus. For each milieu (with the exception of *special groups*) a keyword list was manually created by describing its main characteristics. For triggering marketing campaigns by marketers, an algorithm shall be developed that automatically assigns each contest answer to the most likely target group: we propose the youth milieu as best match for a contest answer, for which the estimated semantic similarity between the associated keyword list and user answer is maximal. In case the highest similarity estimate falls below the 10 percent quantile for the distribution of highest estimates, the special groups milieu is selected (cf. also [29]).

Since the keyword list typically consists of nouns (in the German language capitalized) and the user contest answers might contain a lot of adjectives and verbs as well, which do not match very well to nouns in the W2V vector representation, we actually conduct two comparisons for our W2V based measures, one with the unchanged user contest answers and one by capitalizing every word beforehand. The final similarity estimate is then given as the maximum value of both individual estimates.

The same procedure does not work with CEs, since concepts are always written lower case in MultiNet. To identify corresponding nouns for adjectives and verbs, we use the MultiNet relations CHEA and CHPEA. For instance, (inform.1.1 CHEA information.1.1) or (cold.1.1 CHPA coldness.1.1). These lexical relations are directly stored in the semantic lexicon named HaGenLex. Note that for better illustration, we gave English examples although we investigated German texts.

TABLE II  
MINIMUM AND MAXIMUM AVERAGE INTER-ANNOTATOR AGREEMENTS  
(COHEN'S KAPPA).

Method	Contest		
	1	2	3
Min kappa	0.123	0.295/0.030	0.110/0.101
Max. kappa	0.178	0.345/0.149	0.114/0.209
# Entries	1544	100	100

TABLE III  
ACC. VALUES FOR SEVERAL ESTIMATES. RW: METHOD OF [6] APPLIED  
TO ODENET, CE=SEMANTIC CONCEPT EMBEDDINGS, STV=SKIP  
THOUGHT VECTORS, CE+W2V: WEIGHTED AVERAGING COSINE OVER  
WORD AND CES.

Method	1	2	3	Total
Random	0.152	0.090	0.167	0.146
Jaccard	0.150	0.194	0.045	0.142
GloVe	0.203	0.254	0.303	0.222
RW	0.294	0.254	0.227	0.281
W2V	0.299	<b>0.313</b>	0.258	0.296
STV	0.241	0.194	<b>0.348</b>	0.249
Elmo	0.150	0.224	0.258	0.173
Bert	0.221	0.209	0.212	0.218
CE	0.274	0.299	0.258	0.275
CE+W2V	<b>0.305</b>	0.299	0.303	<b>0.304</b>

## VI. EVALUATION

For evaluation of the marketing group segmentation task, we selected three online contests (language: German), where people elaborated on their favorite travel destination (contest 1), speculated about potential experiences with a pair of fancy sneakers (contest 2), and explained why they emotionally prefer a certain product out of four available candidates. To provide a gold standard, three professional marketers from different youth marketing companies annotated independently the best matching youth milieus for every contest answer. We determined for each annotator individually his/her average inter-annotator agreement with the others (Cohen's kappa). The minimum and maximum of these average agreement values are given in Table II. Since for contest 2 and contest 3, some of the annotators annotated only the first 50 entries (last 50 entries respectively), we specified min/max average kappa values for both parts. We further compared the youth milieus proposed by our unsupervised matching algorithm with the majority votes over the human experts' answers (see Table III) and computed its average inter-annotator agreement with the human annotators (see again Table II).

We evaluated both, our similarity estimate based on CEs alone and its weighted mean with W2V (weight for CE: 0.2, weight for W2V: 0.8).

The W2V, CEs, and Skip Thought Vectors (STV) were trained on the German Wikipedia. The actual document similarity estimation for STVs is accomplished by the usual centroid approach. An issue we are faced with for our evaluation scenario of market segmentation (see Sect. V) is that STVs

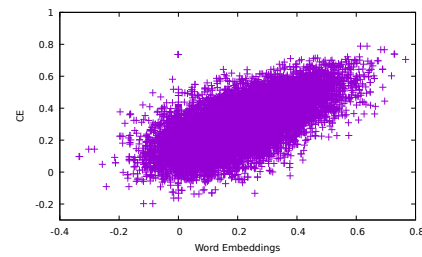


Fig. 2. Scatterplot between similarity estimate based on word embeddings (x-axis) and semantic CEs (y-axis).

are not bag-of-words models but actually take the sequence of the words into account and therefore the obtained similarity estimate between the milieu keyword list and contest answer would depend on the keyword ordering. However, this order could have arbitrarily been chosen by the marketers and might be completely random. A possible solution is to compare the contest answers with all possible permutations of keywords and determine the maximum value over all those comparisons. However, such an approach would be infeasible already for medium keyword list sizes. Therefore, we apply for this scenario a beam search to extend the keyword list iteratively while keeping only the n-best performing permutations.

We also applied the approach of Goikoetxea et al. [6], another embeddings approach based on a linguistic network. Originally, the embeddings have been created by applying random walks on WordNet choosing randomly lexicalizations for the synset nodes. Since WordNet is not available for German and we failed to obtain an academic license for the largest German-based linguistic network GermaNet, we applied their approach to the freely available German linguistic ontology OdeNet [21]. Goikoetxea et al. proposed several combination methods of ordinary W2V and their linguistic network-based word vectors, where the concatenation of both performed best. This is also the approach we evaluated in this paper.

Note that we did not conduct a hyperparameter search on our input parameters like the random walk sentence restart threshold or the 10% quantil for selecting the Special Groups milieu. Instead we selected values based on experience that gave good results in practice. The window size for the embeddings has been chosen with seven rather large to obtain a reasonable number of concepts in the window.

Consider the annotations provided by the marketing experts, the evaluation showed that the inter-annotator agreement values vary strongly for contest 2 part 2 (minimum average annotator agreement according to Cohen's kappa of 0.03 while the maximum is 0.149, see Table II). In general, the kappa values, which estimate inter-annotator agreement are rather small, which insinuates that the manual labeling of the contests and therefore also the automatic labeling process proved to be quite challenging.

In combination with word embeddings, semantic CEs outperforms on our three contests word embeddings alone as

well as the approach of Goikoetxea et al. as well as several deep learning based approaches like German Bert model, multilingual Elmo model, and Skip Thought Vectors. Still, semantic CEs alone perform currently worse than ordinary word embeddings. Sources of errors are the word sense disambiguation (especially the word sense disambiguation of Wocadi) and the lemmatizer LemmaGen, which is currently trained on Wiktionary which provides a good coverage of generic nouns but falls short in terms of named entities. Additionally, Wocadi proved to be quite vulnerable in case of misspelled words or highly complex syntactic structures. In such cases it often only produces a chunk part or the parsing process fails altogether. Figure 2 shows a scatterplot, where we compare the similarity estimate based on word embeddings (x-axis) with its counterpart based on CEs (y-axis). The figure shows that there is a considerable correlation between both estimates. One advantage of our method over using Bert and Elmo embeddings is that the marketer can directly specify the intended word senses as key words for the youth milieus. This is especially important if the keyword list is small and contains not enough textual contexts for Bert and Elmo embeddings to perform well.

## VII. CONCLUSION AND FURTHER WORK

We proposed a novel semantic similarity measure based on semantic CEs. These embeddings were extracted from an SN using random walks on them. All the SNs were automatically generated from Wikipedia using the Wocadi parser. The evaluation showed that semantic CEs outperformed W2V embeddings when used in combination with the latter on the task of assigning participants to the best matching marketing target group. There is still space for further improvement. For instance, inner nodes are currently eliminated from the random walk and therefore not used at all. Instead one could replace them with their associated ontological sort (see Section III).

Currently, our system depends on the Wocadi parser, which is not freely available. Therefore, we aim to test, whether similar results can be obtained, if one uses freely available semantic role labeling parsers instead of Wocadi. A concept representation of each word could hereby be obtained by determining the WordNet / OdeNet / GermNet synset group that best fits to the given context as obtained by a Word Sense Disambiguation.

## ACKNOWLEDGEMENT

Hereby we want to thank Sven Hartrumpf and Sempria GmbH for providing us with semantic network parses from the entire Wikipedia.

## REFERENCES

- [1] S. Hartrumpf, "Hybrid disambiguation in natural language analysis," Ph.D. dissertation, FernUniversität in Hagen, 2002.
- [2] M. Damonte, S. B. Cohen, and G. Satta, "An incremental parser for abstract meaning representation," in *Proceedings of EACL (2017)*, Valencia, Spain, 2017, pp. 536–546.
- [3] C. Fellbaum, *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press, 1998.
- [4] B. Hamp and H. Feldweg, "GermaNet - a lexical-semantic net for German," in *Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, 1997, pp. 9–15.
- [5] T. Trouillon, J. Welbl, S. Riedel, E. Gaussier, and G. Bochar, "Complex embeddings for simple link prediction," in *Proceedings of the 33rd International Conference on Machine Learning*, 2016, pp. 2071–2080.
- [6] J. Goikoetxea, E. Agirre, and A. Soroa, "Single or multiple? Combining word representations independently learned from text and WordNet," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, USA, 2016, pp. 2608–2614.
- [7] A. Kutuzov, A. Panchenko, S. Kohail, M. Dorgham, O. Oliynyk, and C. Biemann, "Learning graph embeddings from wordnet-based similarity measures," *CoRR*, vol. abs/1808.05611, 2018. [Online]. Available: <http://arxiv.org/abs/1808.05611>
- [8] E. L. Menćia, G. de Melo, and J. Nam, "Medical concept embeddings via labeled background corpora," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, Slovenia: European Language Resources Association (ELRA), May 2016, pp. 4629–4636. [Online]. Available: <https://www.aclweb.org/anthology/L16-1733>
- [9] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Katar, 2014.
- [10] A. L. Beam et al., "Clinical concept embeddings learned from massive sources of multimodal medical data," in *Proceeding of the Pacific Symposium of Biocomputing*, 2020, pp. 295–306.
- [11] G.-I. Brokos, Prodromos, and I. Androutsopoulos, "Using centroids of word embeddings and word mover's distance for biomedical document retrieval in question answering," in *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, Berlin, Germany, 2016, pp. 114–118.
- [12] T. Mikolov, I. Sutskever, C. Ilya, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proceedings of the Conference on Neural Information Processing Systems (NIPS)*, Lake Tahoe, Nevada, USA, 2013, pp. 3111–3119.
- [13] E. Gabrilovic and S. Markovitch, "Wikipedia-based semantic interpretation for natural language processing," *Journal of Artificial Intelligence Research*, vol. 34, 2009.
- [14] R. Kiro, Y. Zhu, R. Salakhudinov, R. S. Zemel, A. Torralba, R. Urtasun, and S. Fiedler, "Skip-thought vectors," in *Proceedings of the Conference on Neural Information Processing Systems (NIPS)*, Montréal, Canada, 2015, pp. 3294–3302.
- [15] Y. Song and D. Roth, "Unsupervised sparse vector densification for short text similarity," in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Denver, Colorado, USA, 2015, pp. 1275–1280.
- [16] V. Mijangos, G. Sierra, and A. Montes, "Sentence level matrix representation for document spectral clustering," *Pattern Recognition Letters*, vol. 85, 2017.
- [17] K.-J. Hong, G.-H. Lee, and H.-J. Kom, "Enhanced document clustering using wikipedia-based document representation," in *Proceedings of the 2015 International Conference on Applied System Innovation (ICASI)*, 2015, pp. 183–186.
- [18] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. C. K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proc. of NAACL*, New Orleans, Louisiana, USA, 2018, pp. 2227–2237.
- [19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL*, 2019, pp. 4176–4186.
- [20] E. Alsentzer et al., "Publicly available clinical bert embeddings," 2019.
- [21] M. Siegel and F. Bond, "OdeNet: Compiling a germanwordnet from other resources," in *Global WordNet Conference*, 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:232021541>
- [22] F. Mahdizoltani, J. Biega, and F. M. Suchanek, "YAGO3: A knowledge base from multilingual wikipedias," in *Proceedings of the 7th Biennial Conference on Innovative Data Systems Research (CIDR 2015)*, Asilomar, California, USA, 2015.
- [23] H. Helbig, *Knowledge Representation and the Semantics of Natural Language*. Berlin, Germany: Springer, 2006.
- [24] S. Hartrumpf, H. Helbig, and R. Osswald, "The semantically based computer lexicon HaGenLex – Structure and technological environment," *Traitement automatique des langues*, vol. 44, no. 2, pp. 81–105, 2003.

- [25] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang, "RotatE: Knowledge graph embedding by relational rotation in complex space," in *Proceedings of ICLR*, 2019.
- [26] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yehnenko, "Translating embeddings for modeling multi-relational data," in *Advances in neural information processing systems*, 2013, pp. 2787–2795.
- [27] T. Ebisu and R. Ichise, "TorusE: Knowledge graph embedding on a lie group," in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. AAAI Press, 2018, pp. 1819–1826.
- [28] M. Lynn, "Segmenting and targeting your market: Strategies and limitations," Cornell University, Tech. Rep., 2011, online: <http://scholarship.sha.cornell.edu/articles/243>.
- [29] T. von der Brück and M. Pouly, "Text similarity estimation based on word embeddings and matrix norms for targeted marketing," in *Proceedings of NAACL*, 2019, pp. 1827–1836.