Interpretability of Hungarian embedding spaces using a knowledge base

Vanda Balogh[1], Gábor Berend[1,2], Dimitrios I. Diochnos[3], György Turán[2,4], Richárd Farkas[1]

[1]Szegedi Tudományegyetem, Informatikai Intézet
[2]MTA-SZTE Mesterséges Intelligencia Kutatócsoport
[3]Department of Computer Science, University of Virginia
[4] Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago
{bvanda, berendg, rfarkas}@inf.u-szeged.hu diochnos@virginia.edu gyt@uic.edu

Abstract. While word embeddings have proven to be highly useful in many NLP tasks, they are difficult to interpret for humans. Sparse word embeddings are reminiscent of knowledge bases containing words that are already characterized in sparse forms. In our work, we investigate to what extent sparse word representations convey knowledge about the words in knowledge bases. We utilize Hungarian sparse word embeddings and ConceptNet, a knowledge base that supports Hungarian.
Keywords: sparse word embedding, interpretability, knowledge base, ConceptNet

1 Introduction

Word embeddings generate low dimensional word representations from large corpora. Each word is represented as a vector of real numbers. Word vectors that are similar to each other tend to be semantically related which makes word embeddings effective in a variety of natural language processing tasks. Even though word embeddings perform well in these tasks, they are difficult to interpret for humans. Word embeddings employ dense representations of words, while natural language phenomena are extremely sparse by their nature. Motivated by this sparse behaviour, the construction of word embeddings that were made sparse is getting popular recently [1,2,3,4,5].

Knowledge bases already include words in sparse forms implicitly. The relations in these knowledge bases can describe how the words are related to each other by their lexical definition, and also how they are related through commonsense knowledge. Thus, we have human interpretable features at hand. This appealing characteristic of human assembled knowledge representation has already inspired others to create non-distributional word representations [6].

In our work, we would like to explore the interpretability of the dimensions of distributed word embeddings. As our first experiment, we examine Hungarian sparse embedding matrices by assigning each dimension one concept extracted

from ConceptNet[7], a multilingual knowledge base. One potential application of these assignments can be knowledge graph expanding.

2 Related Work

The explanatory power of distributional semantic models (DSMs) in terms of meaning is not clear as they often provide a quite coarse representation of semantic content [8]. There have been proposals for the semantic evaluation of DSMs, e.g., QVEC[9] and BLESS[10]. The QVEC evaluation measure aims to score the interpretability of word embeddings, a topic close to our research. Dimensions of the word embeddings are aligned with interpretable dimensions – corresponding to linguistic properties extracted from SemCor [11] – to maximize the cumulative correlation of the alignment. BLESS is a dataset designed for the semantic evaluation of DSMs. It contains semantic relations connecting (target and relatum) concepts as tuples. Thus, BLESS allows the evaluation of models by their ability to extract related words given a target concept. The method called the THING RECOGNIZER [12] attempts to make Hungarian embedding spaces interpretable by assigning semantic features to words in a language-independent manner.

In the following, we briefly introduce ConceptNet, a semantic multilingual knowledge base. In our work, we extract interpetable features (concepts) from ConceptNet in order to help exploring the interpretability of the dimensions of word embeddings.

2.1 ConceptNet

Relation	Symmetric	Example Assertion
ANTONYM	✓	deep ↔ shallow
HASCONTEXT	✗	gurl → slang
HASPROPERTY	✗	marsh → muddy and moist
ISA	✗	eagle → bird
MADEOF	✗	ice → water
SYNONYM	✓	bright ↔ sunny
RELATEDTO	✓	torture ↔ pain
USEDFOR	✗	science → understand life

Table 1: Extract of relations from ConceptNet 5.

ConceptNet is a semantic multilingual knowledge base describing general human knowledge collected from a variety of resources including WordNet, Wiktionary and Open Mind Common Sense. ConceptNet can be perceived as a graph whose nodes correspond to words and phrases. The nodes of the semantic network are called *concepts* and the (directed) edges connecting pairs of nodes are called *relations*. The records of the knowledge base are called *assertions*. Each assertion associates two concepts – *start* and *end* nodes – with a relation in the

semantic network and has additional satellite information beyond these three objects; for example, the *dataset* from where the assertion was obtained (e.g., WordNet). Figure 1 provides an example of an assertion found in ConceptNet 5 – the latest iteration of ConceptNet. Relations can be symmetrical, e.g., SYNONYM and RELATEDTO, or asymmetrical, e.g., HASPROPERTY and ISA. An incomplete list of relations present in ConceptNet 5 can be found in Table 1.


```
{
  "dataset": "/d/wikitionary/en",
  "license": "cc:by-sa/4.0",
  "sources": [{"contributor": "/s/resource/wikitionary/en"}],
  "weight": 1.0,
  "uri": "/a/[/r/HasContext/,/c/hu/poligon/n/,/c/en/geometry/]",
  "rel": "/r/HasContext",
  "start": "/c/hu/poligon/n",
  "end": "/c/en/geometry"
}
```


Fig. 1: Example assertion from ConceptNet 5. The *start* and *end* nodes are connected by an edge labelled *rel* corresponding to the relation between the nodes. Assertions feature additional information like *dataset* which represents the source of the assertion and *weight*, the strength of the assertion which is a positive value.

3 Experiments

Our aim is to explore the interpretability of the dimensions of Hungarian embedding matrices by assigning each dimension a human interpretable feature. A somewhat unnatural characteristic of standardly applied word embeddings (e.g. word2vec [13] or Glove [14]) is that the learned vectors have non-zero coefficients everywhere, implying that every word can be characterized with every dimension at least to a tiny extent. From a human perception point of view this dense behavior is quite undesired, because for most features we would not like to see any relation to hold. To approximate the sparse behaviour of natural language phenomena, we employ embeddings that are turned sparse as a post-processing step suggested in [4]. Results from other studies and literature argue that sparse word representations are more interpretable by humans (e.g. word intrusion) and perform well on downstream tasks (e.g. sentiment analysis) [1,3,2,5,15,4].

The rows of a sparse embedding matrix S, correspond to sparse word vectors representing words. We call the columns (dimensions) of the sparse embedding matrix *bases*. As human interpretable features, we take concepts extracted from a semantic knowledge base, ConceptNet, and the sparse embedding we employ is derived from the dense Numberbatch [16] vectors. This way, our goal reformulates to designating a concept to each base.

We basically deal with a tripartite graph (see Figure 2) with words connected to bases – corresponding to the columns of the embedding matrix – and concepts,

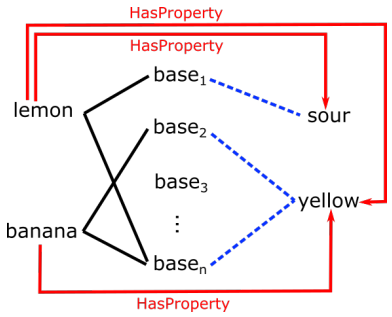

Fig. 2: Tripartite graph presenting the connections between embedded words, bases and concepts. Connections denoted by solid lines are present, our aim is to recover the relations between bases and concepts (dashed lines).

respectively. A word, w is connected to $base_i$ if the ith coordinate of the sparse word vector corresponding to w is nonzero. Also, w is connected to a concept c with label l if there exists an assertion in ConceptNet that associates w and c with the relation l. We are interested in the relations between concepts and bases (dotted lines).

3.1 Hungarian sparse word embeddings

Numberbatch [16] is an embedding approach combining distributional semantics and ConceptNet 5.5 using a variation on retrofitting [17]. The Hungarian sparse word embeddings are derived from dense Numberbatch embeddings related to Hungarian concepts (i.e., concepts prefixed with `/c/hu/`). As a side note, the words present in ConceptNet align much better with the vocabulary provided by Numberbatch than with other embeddings' vocabulary. This is because Numberbatch implicitly makes use of words (and their specific forms) from ConceptNet and any arbitrary embedding would include a vast amount of forms of a single word since Hungarian is a morphologically rich language.

Sparse embeddings $\mathbf{s}_i$ are derived from dense embeddings $\mathbf{x}_i$ according to the objective function

$$\min_{D\in\mathcal{C},s} \frac{1}{2n}\sum_{i=1}^{n}\left(\|\mathbf{x}_i - D\mathbf{s}_i\|_2^2 + \lambda\|\mathbf{s}_i\|_1\right),$$

where D is a dictionary matrix of basis vectors with length not exceeding 1. The regularization constant, λ controls the sparsity of the resulting embeddings s_i. As λ increases, the density of the nonzero coefficients in s_i decreases. In total, we use four sparsity levels according to λs from $\{0.2, 0.3, 0.4, 0.5\}$. Table 2 shows sparsity of each sparse embedding matrix. We have a vocabulary of 17k words which are embedded into a vector space of 1000 dimensions.

λ	0.2	0.3	0.4	0.5
sparsity	99.66%	99.81%	99.88%	99.92%

Table 2: The ratio of zero elements to all the elements from sparse embedding matrices with λ regularization coefficient.

3.2 Hungarian ConceptNet

We utilize the Hungarian part of ConceptNet 5.5 and ConceptNet 5.6. in our experiments. Every assertion has a *start* and *end* node, which are connected by a directed labeled edge where the label is specified by the relation between the nodes. Basically, an assertion is a triplet of (start node, relation, end node). If the relation is symmetric, the connecting edge is bidirectional. In the following, we refer to start nodes as (embedded) *words* and end nodes are regarded as *concepts* (which should not be confused with the concepts mentioned in Section 2.1). These end nodes – seen as concepts – will be assigned to the bases of the sparse embedding matrix.

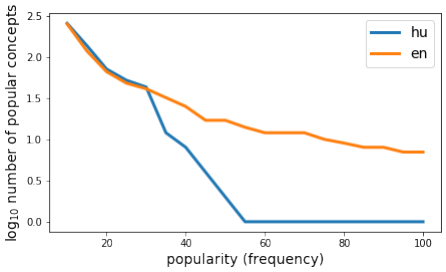

Fig. 3: Comparison on the number of English and Hungarian concepts that appear frequently (above frequency) as end nodes in assertions. The y axis shows the $\log_{10}$ of the number of concepts that are frequent.

For our experiments, we produce the subgraph of ConceptNet which encodes useful information on Hungarian (embedded) words. It is important to note that English is a core language of ConceptNet (i.e. the language is admittedly well supported) while Hungarian is not. First, we take the assertions associating two Hungarian nodes and to further diversify them, we adopt assertions associating a Hungarian start node with an English end node. It is worth to expand the set of concepts with English concepts, because there are significantly more English concepts that appear a lot as end nodes of assertions i.e. among the popular concepts there are more English ones (see Figure 3). To avoid redundancy, the assertions including symmetric relations that connect two Hungarian concepts are dropped. Instead, these groups of Hungarian words defined by symmetric relations (eg. synsets via the Synonym relation) are represented by English end nodes. In other words, the assertions including symmetric relations between Hungarian and English are kept in order to group together Hungarian concepts connected by symmetric relations according to their English equivalent. As an

example the Hungarian synonyms "ronda", "csúnya" and "ocsmány" are all connected to the English "ugly" through a SYNONYM relation. Instead of working with the complete graph of these Hungarian words that contains unnecessary information, we simply make use of the information that they can be grouped together by the English "ugly".

assertion \ version	**ConceptNet 5.5**	**ConceptNet 5.6**
any → any	28 million	32 million
hu → hu	31984	51819
hu → en	57941	61666
hu → (hu ∨ en)	89925	113485
filtered hu → hu	23844	42403
filtered hu → (hu ∨ en)	81785	104069

Table 3: Summary on the number of assertions in ConceptNet 5.5 and ConceptNet 5.6. The assertion types are listed according to the languages of the connected nodes. The filtered assertions disregard possible assertions associating two Hungarian nodes with a symmetric relation.

Altogether, we have a result of 81k and 104k assertions from ConceptNet 5.5 and 5.6, respectively. Further on, we will refer to the resulting subsets of ConceptNet globally as *Hungarian Conceptnet* (HCN) and use it in our experiments. The version 5.5 or 5.6 (of HCN) is always specified if required. Table 3 summarizes the number of assertions present in HCN 5.5 and 5.6. For further purposes, we experiment with end nodes that are with the connecting relation to reflect the meaning of assertions. We call this approach *augmented* in terms of the representation of end nodes – seen as concepts. So the assertion associating "eb" and "dog" with the SYNONYM relation has its start node "eb" and its end node is "dog/SYNONYM".

Basically, HCN 5.5 is used for association of concepts to bases and HCN 5.6 is used for evaluation. All in all, there are 48k and 58k distinct end nodes included in HCN 5.5 and HCN 5.6., respectively. If we ignore the relations by which end nodes were augmented we get 44k and 53k different relations. Although relations may be important in terms of meaning, we may resort to ignoring them to be able to further group together words according to their connecting concepts. Ignoring relations is also motivated by assertions like (a_fiók, SYNONYM, jacket), which presents a case where probably the relation SYNONYM is wrong between the word "a_fiók" and "jacket". A relation like ATLOCATION or RELATEDTO would fit better. In general, we use two types representation for concepts (ends nodes): the ones ignoring relations and the augmented approach.

Overall, there are 26 types of relations present in HCN. Surprisingly, there are relations for which there are substantially fewer assertions in HCN 5.6 than HCN 5.5 (see Figure 4). A lot of relations have the same number of assertions in both versions of HCN. The relation ETYMOLOGICALLYDERIVEDFROM is not present in HCN 5.5 and some of the richest relations include DERIVEDFROM, FORMOF, RELATEDTO and SYNONYM.

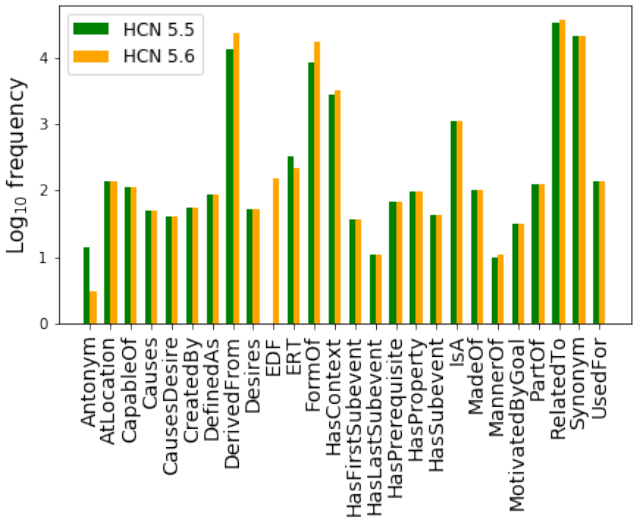

Fig. 4: Log_{10} frequency of the assertions in HCN 5.5 and 5.6 according to relations. The relation EDF is short for EtymologicallyDerivedFrom and ERT refers to EtymologicallyRefersTo.

3.3 Phases of association

The process of associating a base with a concept is divided into four phases. First, we produce an adjacency matrix based on HCN 5.5, then we multiply its transpose with the sparse word embedding matrix. Afterwards, the positive pointwise mutual information (PPMI) values of the resulting matrix are computed and finally, the association takes place by taking the argmax of the matrix containing PPMI values. The four phases are detailed below. Figure 5 provides an overview of the four phases.

I. Produce ConceptNet matrix. Given HCN 5.5 (described in §3.2), we consider it as a bipartite graph whose two sets of vertices correspond to two ordered sets containing the start and end nodes of assertions, respectively. The start nodes are regarded as (possibly embedded) words and the end nodes as concepts. The bipartite graph is represented as a biadjacency matrix C (which simply discards the redundant parts of a bipartite graph's adjacency matrix). Every word w corresponding to a start node is associated with an indicator vector v_w where the ith coordinate of v_w is 1 if w is associated to the ith end node, 0 otherwise. At this point, words can have two sparse representations: the vectors coming from sparse word embeddings and the binary vectors provided by HCN. To differentiate them, we call the former ones *embedded vectors* and the latter ones *ConceptNet vectors*. It is important to note, that it is possible for an embedded word to lack its ConceptNet vector representation if the word itself is not present in the set of start nodes. On another note, there are words

in HCN 5.5 that are not presented in the vocabulary of the embedding; that is, they do not have embedded vector representations.

II. Compute product. We binarize the nonnegative sparse embedding matrix S by thresholding it at 0, then we take the product of the transpose of C and the binarized version of S. The result is a dense matrix A, whose element at the ith row and jth column equals the number of words the ith concept and the jth base (from the sparse embedding matrix) appear together.

III. Compute PPMI. To generate a sparse matrix from the dense A matrix, we compute its positive pointwise mutual information (PPMI) for every element. PPMI for the ith concept c_i and jth base b_j is computed as

$$\mathrm{PPMI}(c_i, b_j) = \max\left(0, \ln \frac{\mathrm{P}(c_i, b_j)}{\mathrm{P}(c_i)\mathrm{P}(b_j)}\right),$$

where probabilities are approximated as relative frequencies of words as follows: $\mathrm{P}(c_i)$ is the relative frequency of words connected to the ith concept, $\mathrm{P}(b_j)$ takes the relative frequency of words whose jth coefficient in their embedded vector representation is nonzero and $\mathrm{P}(c_i, b_j)$ is the relative frequency of the co-occurrences of the words mentioned above. The result is a sparse matrix P whose columns correspond to bases, and its rows correspond to concepts.

IV. Take argmax. By taking the arguments of the maximum vales of every column in P we can associate a base with a concept.

```
Association(sparse_word_embedding, conceptnet){
    nodes = {(start, end) in conceptnet}
    C = biadjacency(nodes)
    A = transpose(C) * binarize(sparse_word_embedding)
    P = PPMI(A)
    max_concepts = argmax(P, max_by=columns)
    //the ith element is the concept associated with the ith base
    return max_concepts
}
```

Fig. 5: The process of associating concepts to bases summarized in pseudocode.

4 Evaluation metrics

To evaluate the associations between bases and concepts, we employ HCN 5.6. We are interested if the new assertions compared to HCN 5.5 are presented in the associations (which can be perceived as a link prediction task). We would like to measure if the prominent words of the ith base (i.e., the words whose ith embedded coordinate is nonzero) are in relation with the concept associated

to the ith base according to HCN 5.6 (only new assertions are considered). We define the set D_{b_i} as the set that contains the prominent words of base b_i, i.e.

$$D_{b_i} = \{w_j | w_j(i) > 0, 1 \leq j \leq n\}$$

where $w_j(i)$ is the ith coordinate in the embedded vector representation of w_j and n is the size of the vocabulary of the sparse embedding. It is worth to mention that D_{b_i} only depends on the embedding itself. Furthermore, we define F_{b_i} as the subset of D_{b_i} that contains words which are present in the new assertions as start nodes and have a connecting end node to the concept that is associated to b_i, formally

$$F_{b_i} = \{w_j | w_j \in D_{b_i} \text{ and } (w_j, \mathit{concept}(b_j)) \in N\},$$

where $\mathit{concept}(b_j)$ is the concept associated to b_i and N contains *(start node, end node)* pairs that make a new assertion to HCN 5.6. The following information retrieval measures are used for evaluation:

Mean Reciprocal Rank The reciprocal rank (RR) of the ith base, b_i is

$$\mathrm{RR}(b_i) = \frac{1}{\mathrm{rank}(w_{F_{b_i}})},$$

where $w_{F_{b_i}}$ is the word from F_{b_i} with the highest coefficient in b_i, and for a word, w, $\mathrm{rank}(w)$ is the rank of w among D_{b_i}, so that the word with the largest coefficient in D_{b_i} has a rank of 1, and the word with the smallest coefficient has a rank of $|D_{b_i}|$. Mean Reciprocal Rank (MRR) is the mean of the reciprocal ranks over all the bases.

Mean Precision The precision of base b_i is computed as

$$\mathrm{Prec}(b_i) = \frac{|F_{b_i}|}{|D_{b_i}|}.$$

This way, mean precision (MR) equals $\frac{1}{m}\sum_{i=1}^{m} \mathrm{Prec}(b_i)$, where m is the number of bases.

Mean Average Precision The average precision of base b_i is the following:

$$\mathrm{AvgPrec}(b_i) = \frac{1}{|D_{b_i}|}\sum_{k=1}^{n} \frac{|F_{b_i}^k|}{|D_{b_i}^k|},$$

where both $F_{b_i}^k$ and $D_{b_i}^k$ are cutoffs of F_{b_i} and D_{b_i}, respectively, so that the words are restricted to the first k words coming from the embedding vocabulary (of size n). Mean average precision is the mean of average precisions over the bases.

In addition to all the above, we examine these metrics in the light of all the assertions in HCN 5.5 and 5.6. In this case we only have to alter the definition of the set F_{b_i} to

$$\hat{F_{b_i}} = \{w_j | w_j \in D_{b_i} \text{ and } (w_j, \mathit{concept}(b_j)) \in \hat{N}\},$$

where $\hat{N}$ contains *(start node, end node)* pairs that make an assertion in HCN 5.5 or 5.6, i.e., we let the assertions of both HCN versions to overlap.

5 Results and Discussion

Based on PPMI values, we map a concept to each base. Some associations can be inspected in Table 4. The first thing we notice is that English concepts are much more frequent than Hungarian ones. The proportion of English concepts is ranging between 98.9% and 100% for all λs. One reason behind this is definitely that within HCN 5.5 the number of different English concepts is 35k (38k augmented), while the number of Hungarian concepts is 9k (9k augmented). Also, English concepts are more popular (see Figure 3).

Some concepts associated with bases reflect the dominant words of the bases (the words that had the highest coefficient in the specific base). However, some concepts seem to have nothing in common with the dominant words of the associated base. This might be because some of the less dominant words of the base contribute to the PPMI. For example, the concept "en/accident/Synonym" is associated with the 10th base of the sparse embedding matrix (λ=0.2) whose most dominant words include "személygépkocsi", "automobil", "autós", "kocsi", "autó", however, there are words – like "baleset", "mentőautó", "gázol", "ütközés", "gyorshajtás", "ittas vezetés" – with smaller coefficients in the base that have more in common with the concept itself.

base	concept	PPMI	most dominant words of base
875	en/dehydrated/RelatedTo	7.978	aszalt szilva, dunyha, birsalma, birs, birskörte
533	en/hard_disk/RelatedTo	7.824	szerves, szervesen, farész, merisztéma, háncsrész
927	en/audacity/RelatedTo	7.690	habar, malter, habarcs, vakolat, mozsár
243	en/beach/Synonym	7.285	mamusz, pacsker, papucs, vietnami papucs, strandpapucs
593	en/adult/RelatedTo	7.285	érett, andragógia, felnő, felnőtt, felnőttképzés
327	en/absinthe/RelatedTo	4.549	odavisz, odaad,idead, átad, ad
15	en/about_cardinality/HasContext	4.481	irracionális szám, háromszögszám, numerikus analízis, másodfokú függvény, logaritmusfüggvény
431	en/about_cardinality/HasContext	4.413	ezik, eszt, aszt, lél, lál
709	en/agape/Synonym	4.248	többé kevésbé, a volánnál, mihelyt, dagály, egymás
814	en/abscess/Synonym	3.856	spongyabob kockanadrág, szemérem, rosszban, sárgavállú amazon, flerovium

Table 4: Association pairs scoring the highest and lowest PPMI values at sparse embedding with $\lambda = 0.2$ using concepts augmented with relation type.

The associations are evaluated on four sparse word embeddings (based on their regularization constants) with two types of concept sets (with or without relations). Three evaluation measures connected to information retrieval (MRR,

MP, MAP) are used which can focus on either the new assertions to HCN 5.6 or all the assertions present in HCN 5.5 and 5.6. Table 5 shows the evaluation scores. The results of new assertions to HCN 5.6 naturally have very low scores. This is because, out of the 48k concepts in HCN 5.5, at most 1k is used for the associations and most of the concepts (end nodes) present in the new assertions come from the remaining 47k concepts. On average, there is only 35 common concepts between the concepts coming from the associations and the new assertions to HCN 5.6. Also, we know that there are around 20k more assertions in HCN 5.6, however this version of HCN introduces 5k assertions with concepts (end nodes) not present in HCN 5.5.

λ	aug	MRR	MP	MAP
0.2	✗	0.00009	0	0
	✓	0.00026	0	0
0.3	✗	0.00117	0	0
	✓	0.00126	0	0
0.4	✗	0.00013	0	0
	✓	0.00121	0	0
0.5	✗	0.00113	0.00021	0.00005
	✓	0.00165	0.00017	0.00003

(a) Evaluation in terms of new assertions to HCN 5.6.

λ	aug	MRR	MP	MAP
0.2	✗	0.02093	0.00333	0.00487
	✓	0.02197	0.00334	0.00485
0.3	✗	0.03780	0.00812	0.01178
	✓	0.04086	0.00829	0.01228
0.4	✗	0.05440	0.01559	0.01858
	✓	0.05661	0.01553	0.01963
0.5	✗	0.06731	0.02797	0.03090
	✓	0.06917	0.02573	0.02960

(b) Evaluation on all assertions available to HCN 5.5 and 5.6

Table 5: Evaluation scores for associations using sparse embeddings with λ regularization constant, and concepts from HCN 5.5 either including relations or ignoring them.

We can observe that sparser embeddings (with higher λ values) perform significantly better in terms of all evaluation metrics. Moreover, embeddings with lower λ values miss most of the new assertions to HCN 5.6, which results in near zero scores. A reason for that can be that "less sparse" embeddings contain too much noise. Ignoring relations of concepts definitely helps the new assertions in terms of MP and MAP, although this is not true for all the assertions, generally. However, the MRR values are consistently better at augmented representations of concepts: on average the 16th most dominant word of each base (of the sparse embedding with $\lambda = 0.5$) is connected to the associated concept in either HCN5.5 or 5.6. The augmented representation of concepts restricts associations, but gives more precise results.

Some highlights from the best performing associations can be seen in Table 6. We can notice that some of the dominant words of specific bases do not actually form assertions in HCN together with the associated concepts. Thus, the power of the resulting associations resides in the ability to augment existing knowledge bases or improve their quality.

base	concept	dominant words from base
46	en/association_football/RelatedTo	**labdarúgó**, **futball**, labdarúgás, **foci**, **focilabda**
198	en/athletics/HasContext	**súlygolyó**, **súlyemelés**, **súlyemelő**, **súlyú**, súly
773	en/dustman/RelatedTo	szemetes, **hulladék**, **szemeteskonténer**, **szemét**, **szemetet**
139	en/bespectacled/RelatedTo	**optika**, **optikus**, **lencse**, **szemlencse**, **szemüveg**
43	en/building_material/HasContext	**kőkemény**, **mészkő**, **kőbánya**, **homokkő**, **kő**

Table 6: Example for remarkable associations between concepts and bases with some of the dominant words in the base. Words in **bold** do not form an assertion in HCN with the associated concept, making the resulting associations applicable in knowledge base expanding.

6 Conclusion

The general theme of our study was the interpretability of Hungarian word embeddings. We experimented with associating a property (concept) for each column of the embedding matrix. Motivated by the sparse behaviour language phenomena, we employed sparse word embeddings which provide sparse vectorial representation for words.

We utilized four Hungarian sparse embeddings provided by Numberbatch. The concepts were extracted from ConceptNet 5 with Hungarian language in focus. English concepts were adopted to enrich the set of concepts and because they were more popular among assertions. The concepts could either be augmented with relations or not. The strategy of association was based on PPMI values. To measure how the associations reflect the assertions, we introduced four metrics, namely Mean Reciprocal Rank, Mean Average Reciprocal Rank, Mean Precision and Mean Average Precision.

Overall, the results confirm the results of [10,9] that word embeddings have semantic content related to word meaning, and provide a further step towards identifying such word meanings explicitly. We can conclude that sparser representations (with higher λ) perform better in terms of all evaluation metrics, probably because they contain less noise. The augmented approach of concept representations seems to have more precise results in terms of ranking, but it is subject to noise. On the other hand, the non-augmented approach may be more comprehensive. Also, the results indicate that associations may provide help in expanding knowledge bases, especially ConceptNet. In our ongoing work we explore more general forms of word meaning.

Acknowledgments

This work was supported by the National Research, Development and Innovation Office of Hungary through the Artificial Intelligence National Excellence Program (grant no.: 2018-1.2.1-NKP-2018-00008).

References

1. Subramanian, A., Pruthi, D., Jhamtani, H., Berg-Kirkpatrick, T., Hovy, E.H.: SPINE: sparse interpretable neural embeddings. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018. (2018) 4921–4928
2. Murphy, B., Talukdar, P.P., Mitchell, T.M.: Learning effective and interpretable semantic models using non-negative sparse embedding. In Kay, M., Boitet, C., eds.: COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, 8-15 December 2012, Mumbai, India, Indian Institute of Technology Bombay (2012) 1933–1950
3. Faruqui, M., Tsvetkov, Y., Yogatama, D., Dyer, C., Smith, N.A.: Sparse overcomplete word vector representations. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics (2015) 1491–1500
4. Berend, G.: Sparse coding of neural word embeddings for multilingual sequence labeling. Transactions of the Association for Computational Linguistics **5** (2017) 247–261
5. Sun, F., Guo, J., Lan, Y., Xu, J., Cheng, X.: Sparse word embeddings using l1 regularized online learning. In: IJCAI, IJCAI/AAAI Press (2016) 2915–2921
6. Faruqui, M., Dyer, C.: Non-distributional word vector representations. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Association for Computational Linguistics (2015) 464–469
7. Speer, R., Havasi, C.: Representing general relational knowledge in conceptnet 5. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), European Language Resources Association (ELRA) (2012)
8. Lenci, A.: Distributional models of word meaning. Annual Review of Linguistics **4**(1) (2018) 151–171
9. Tsvetkov, Y., Faruqui, M., Ling, W., Lample, G., Dyer, C.: Evaluation of word vector representations by subspace alignment. In: EMNLP, The Association for Computational Linguistics (2015) 2049–2054
10. Baroni, M., Lenci, A.: How we BLESSed distributional semantic evaluation. In: Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics. GEMS '11, Stroudsburg, PA, USA, Association for Computational Linguistics (2011) 1–10
11. Miller, G.A., Leacock, C., Tengi, R., Bunker, R.: A semantic concordance. In: HLT, Morgan Kaufmann (1993)

12. Novák, A., Novák, B.: Cross-lingual generation and evaluation of a wide-coverage lexical semantic resource. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018), European Language Resource Association (2018) 45–51
13. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. CoRR **abs/1301.3781** (2013)
14. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, Association for Computational Linguistics (2014) 1532–1543
15. Vyas, Y., Carpuat, M.: Sparse bilingual word representations for cross-lingual lexical entailment. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, California, Association for Computational Linguistics (2016) 1187–1197
16. Speer, R., Chin, J., Havasi, C.: Conceptnet 5.5: An open multilingual graph of general knowledge (2017)
17. Faruqui, M., Dodge, J., Jauhar, S.K., Dyer, C., Hovy, E., Smith, N.A.: Retrofitting word vectors to semantic lexicons. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics (2015) 1606–1615