Chaining Polysemous Senses for Evocation Recognition

Arkadiusz Janz and Marek Maziarz(✉)

Wroclaw University of Science and Technology, 50-370 Wrocław, Poland
{arkadiusz.janz,marek.maziarz}@pwr.edu.pl

Abstract. In this paper we present a new individual measure for the task of evocation strength prediction. The proposed solution is based on Dijkstra's distances calculated on the WordNet graph expanded with polysemy relations. The polysemy network was constructed using chaining procedure executed on individual word senses of polysemous lemmas. We show that the shape of polysemy associations between WordNet senses has a positive impact on evocation strength prediction and the measure itself could be successfully reused in more complex ML frameworks.

Keywords: Evocation strength · Graph-based similarity · WordNet · Semantic similarity · NLP

1 Introduction

A computational linguist looking for a good description of lexico-semantic subsystem will reach out to electronic dictionaries and thesauri. Princeton WordNet is a prominent example of such a computational model of mental lexicon [7]. Unfortunately, the knowledge obtainable from a dictionary is not sufficient to predict all possible sense relations, especially the associations useful in evocation recognition. Evocations are simply associations of meanings [3]. These semantic couplings go across different parts of speech and jump from one semantic category to another [15].

The fact that probably affects evocation recognition is the absence of lexico-semantic resources other than the ones built up mainly from taxonomic relations (hyponymy, meronymy etc.). Polysemy is yet another type of semantic relatedness that could facilitate noticing some hidden associations. It links distant parts of our lexicon through lexicalised metaphor or metonymy.

Lexical polysemy is a linguistic phenomenon characterised by the coexistence of two or more semantically related senses tied to the same lemma [14]. Some words are monosemous and they have only one meaning, e.g. *smartphone* or *lexical*. Polysemous words can express multiple meanings, e.g. *castle* ('fortified building', 'imposing old mansion', 'rock in chess') or *line* (with dozens of meanings). In the actual usage, words may slightly change their (basic) meanings adjusting to a particular sentence context. This introduces a myriad of contextually motivated meaning shades.

N. T. Nguyen et al. (Eds.): ICCCI 2020, LNAI 12496, pp. 792–804, 2020.
https://doi.org/10.1007/978-3-030-63007-2_62

In many theories semantic nets are used as models of polysemy. Especially, relational semantics treats related polysemy senses as a kind of semantic net [10] often called *polysemy net* (cf. [19]). One particular kind of lexical net is of great importance to linguists: wordnets. Wordnets have been widely used in experiments regarding polysemy, e.g. [1,9]. In this paper we focus on constructing a high-quality lexical resource for English based on Princeton WordNet and polysemy links, for the needs of recognizing evocation strength. As far as we know, earlier studies ignored the properties of polysemy nets that could be a source of useful semantic knowledge.

2 Related Work

Evocation Data Sets. In this paper we use a gold standard evocation data set [3] that contains a list of paired senses with manually assigned scores representing association (evocation) strength. Around 120,000 concept pairs were randomly selected from the set of 1,000 WordNet core synsets. As the data was annotated by multiple annotators, we averaged the scores assigned to evocation pairs, which is a standard procedure in the treatment of this resource (cf. [11]).

Wordnets lack the full description of polysemy. Senses are simply enlisted within different synsets and then integrated with the whole net of paradigmatic relations (like hyponymy or antonymy), and they remain unrelated, unless they are linked via taxonomic relations (like auto-hyponymy). On the other hand, wordnets do contain derivational relations. Why do they then omit polysemy links representing *semantic* derivation? If one considers this problem, they immediately discover that this asymmetry is unjustified [13, p. 120, 183].

This weakness of wordnets cannot be easily resolved, since it is not obvious which senses should be linked together, mainly because of the difficulty in distinguishing between homonymous and polysemous pairs. Undoubtedly, this situation affects evocation recognition. If we want to efficiently match associated senses, we should have as accurate model of mental lexicon as possible.

Evocation Recognition. Evocation recognition is considered a difficult task for NLP. It is said to be more difficult than similarity/relatedness recognition, since something more than bare taxonomy structure is needed to sufficiently predict the evocation strength [6]. Simple WordNet-based similarity measures are very inefficient in recognizing the strength of evocation [3].

The recent progress in distributional modelling and knowledge-based embeddings allowed to design more effective approaches for many different NLP tasks, and some of them were directly applied to the task of evocation recognition. The model proposed by Hayashi [11] combined many different features, e.g. wordnet-based similarity and relatedness features, lexical resource features, and distributional features. Surprisingly, most of them were of the same significance. He argued that progress in this area could be obtained only by introducing to a computational model a new high-quality individual measure.

Researchers working on similar tasks, i.e., *word* associations [4,12], reported the same conclusions. Knowledge-based similarity measures were performing slightly worse, and the best way to achieve higher correlations in the task was to combine many different individual measures [4].

Polysemy Chains. Polysemy topologies were studied by Ramiro et al. [17] in the context of English language evolution. Starting from the etymologically first sense, the authors were able to successfully reproduce the order in which senses appeared during the millennia of history of the English language. They studied several net construction algorithms: inter alia – random, prototype, progenitor and nearest-neighbor ones. In their experiments the latter one achieved the best results.

3 Polysemy Nets in Evocation Recognition

In our approach a graph for evocation recognition combines Princeton WordNet with polysemous links extracted from WordNet glosses (sense definitions).[1] We decided to expand the base graph with three different polysemy networks, each of which might have been treated as a graph model of real lexical polysemy. We tested the following structures:

- A complete polysemy graph which – for a given lemma – linked all its senses together ("WN-g-co"). The graph was built for each polysemous lemma existing in the graph.
- An incomplete graph built by extracting polysemy links from SemCor [5]. We constructed the second graph out of those sense pairs that closely co-occurred in the same text (symbol "WN-g-sc").
- The last model was more sophisticated, as we tried to predict contemporary semantic relations between senses of all polysemous words/lemmas on the basis of WordNet structure. We used here the nearest-neighbor chaining algorithm ("WN-g-ch", Sect. 3.1).

With the graphs we proceeded in the following way:

- We optimised edge weights for each polysemy network type and chose the best polysemy model (Sect. 3.1).
- We tested the impact of a selected model on the evocation strength recognition within a ML framework (Random Forest and Multi-layer Perception, Sect. 3.2).

To avoid overfitting the evocation data set was divided into three parts. A subset of 2,000 evocation pairs was devoted to setting optimal weights in each polysemy graph (*evo2k* set). The next subset of 10,000 evocation instances was used to evaluate the performance of each model in predicting evocation strength, as well as to perform attribute selection for the Random Forest framework and for setting the best Neural Network topology (*evo10k* set). The remaining 108,000 synset pairs (*evo108k* set) were used as a test set to evaluate our model and compare the results with [11].

[1] WordNet glosses were semi-automatically interlinked with contextually appropriate synsets, https://wordnetcode.princeton.edu/glosstag.shtml.

The Complete Polysemy Graph was constructed out of the list of all WordNet senses for each polysemous lemma. For a given m-sense lemma l we linked all its senses, obtaining $\frac{m\cdot(m-1)}{2}$ bidirectional sense relations.

The SemCor Polysemy Graph. SemCor is a sense annotated sub-part of the Brown Corpus [5]. The annotations were based on WordNet. We decided to use this language resource assuming that senses occurring in the same context must be semantically related. This assumption allowed us to retrieve distributional properties of word meanings and map them onto WordNet graph. As a dominant direction of each link, we chose the one from the consecutive meaning to the preceding one, hence in the sequence of sense occurrences $(s^t_{l,1}, s^t_{l,2}, s^t_{l,3}, ...)$ of the same lemma l taken from the text t we established a polysemy relation $s^t_{l,i} \rightarrow\ s^t_{l,(i-1)}$ between neighbouring word occurrences, with an additional constraint that inequality $s^t_{l,i} \neq\ s^t_{l,(i-1)}$ was fulfilled (i.e., only different senses of the same lemma were linked).

3.1 Nearest-Neighbor Chaining Algorithm

To evaluate the impact of polysemy nets on evocation strength recognition we decided to introduce polysemy links between WordNet senses. This task is not trivial, since the actual shape of polysemy associations still remains mysterious for the present day linguistics.

Let us consider three different polysemy net topologies tested in this paper. Figure 1 presents polysemy nets for the word *slaughter*. A complete graph simply links all senses together. SemCor-based polysemy net just groups such sense pairs that co-occur in the corpus, giving rather poor completeness but probably good precision. The chaining algorithm tries to connect senses that are the closest in the WordNet graph. The difficulty of the task is demonstrated by a polysemy net constructed manually on the basis of dictionary descriptions (based on three contemporary English dictionaries, namely – Oxford Lexico[2], Merriam-Webster[3] and Cambridge Dictionary[4], and on the etymological English dictionary[5]).

In this paper we adopted the nearest-neighbor approach as presented in [17] to construct polysemy nets. Ramiro et al. [17] tried to predict the order of appearance of different word senses in the history of English starting with one sense given by an oracle (from a historical English dictionary). They found that the best results were obtained with the use of the chaining algorithm. We applied their algorithm with two main modifications:

- Since we are not aware which sense should be fixed as the first one, we try to deduce it from vertex degrees.
- We apply an asymmetric measure of distance in the directed WordNet+glosses graph (*WN-g*).

[2] https://www.lexico.com/.
[3] https://www.merriam-webster.com/.
[4] https://dictionary.cambridge.org/.
[5] https://www.etymonline.com/.

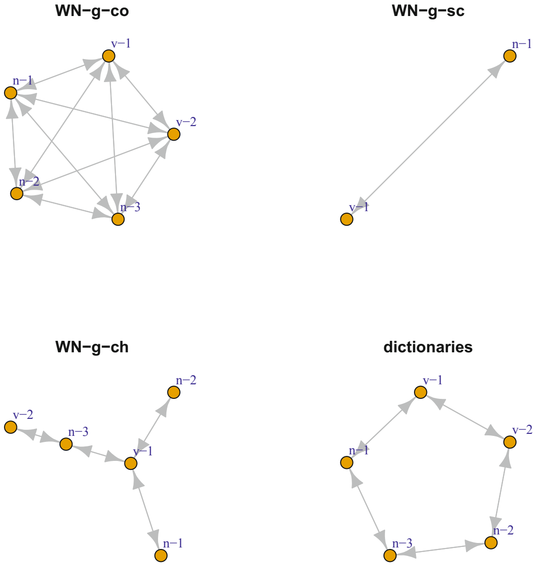

Fig. 1. Three polysemy net topologies for the word *slaughter*: complete graph *WN-g-co* on the topmost left, SemCor-based *WN-g-sc*, and the network built with nearest neighbor chaining algorithm *WN-g-ch*. In the bottom-right corner we present a polysemy net extracted manually from four English dictionaries (three contemporary and an etymological one), taking into account direct proximity of senses in *slaughter* entries of the dictionaries. WordNet definitions are as follows: n-1 – '*the killing of animals (as for food)*', n-2 – '*a sound defeat*', n-3 – '*the savage and excessive killing of many people*', v-1 – '*to kill (animals) usually for food consumption*', v-2 – '*to kill a large number of people indiscriminately*'.

The First Sense Choice. We start with computing the importance scores for each node (synset) based on a vertex degree measure. To be more specific, we calculate the vertex degree $deg(v)$ being the harmonic mean of two different vertex degree measures – the square root of the number of edge instances $\sqrt{deg_i(v)}$ and the number of edge types $deg_t(v)$:

$$deg(v) = \frac{2 \times \sqrt{deg_i(v)} \times deg_t(v)}{\sqrt{deg_i(v)} + deg_t(v)}. \tag{1}$$

The chaining algorithm starts from the node with the highest $deg(v)$ measure.

Geodesics in Polysemy Nets. For each lemma we compute the shortest paths in the directed WordNet graph (*WN-g*) between its senses (synsets). We treat the length of the shortest path as an asymmetric measure of a distance between graph vertices (synsets). We denoted this with $Dist(v_1, v_2)$.

The Chaining Algorithm. Let us assume that we have an m-sense lemma. The chaining algorithm proceeds in the following way:

- Step 1: We start with establishing the first sense that has the highest vertex degree value $deg(v)$. We call such a vertex *fixed*, i.e., $v_j = v_j^{fix}$ iff $deg(v_j) = \max[deg(v_i)]$, where $i \in I$, I being the set of non-fixed vertices, $j \in F$, F being the set of fixed vertices.
- Step 2: For each remaining vertex $v_i, i \in I$, we check the distances $Dist(v_i, v_j^{fix})$ to all fixed vertices $v_j^{fix}, j \in F$, and establish the edge $v_l \rightarrow v_k^{fix}$, iff $Dist(v_l, v_k^{fix}) = \min[Dist(v_i, v_j^{fix})]$, $i, l \in I, j, k \in F$. Again, we call the newly attached l^{th} vertex *fixed*, i.e., $v_l = v_l^{fix}$, $l \in F$.
- We repeat Step 2 in a loop until all vertices are fixed. We call the set of edges $\{v_i^{fix} \rightarrow v_j^{fix}\}_{i \neq j}$ the *polysemy net*, where $i, j \in F = \{1, 2, ..., m\}$. At the end $I = \emptyset$.

Dijkstra's Distance in Modified Nets. We use the polysemy nets to expand WordNet graph. On such modified graph we compute the semantic distance between concepts (with all edge weights set as 1) using Dijkstra's shortest path algorithm. Out of the new distance measure $dist_{Dijkstra}$ we construct the final evocation measure $DSch$:

$$DSch = \begin{cases} \frac{1}{dist_{Dijkstra}} & , when \quad dist_{Dijkstra} \geq 1 \\ 1 & , when \quad dist_{Dijkstra} < 1 \end{cases} \tag{2}$$

For each synset in the evocation set we calculated the $dist_{Dijkstra}$ measure and compared $DSch$ to the evocation strength. The performance of a proposed similarity measure was evaluated with the use of Spearman's correlation ρ.

Optimisation. We tested the impact of different cost values of newly introduced polysemy links on evocation strength prediction. The cost values were used as weights for Dijkstra's shortest path algorithm. For *WN-g-co* and *WN-g-sc* models the weights in the graphs were equal to 1.0, then they were multiplied by optimisation parameters (marked with capital letters A or B). In the case of *WN-g-ch* graph, we took the shortest path length (the geodesic) in *WN-g* as a base cost (as described above), the basic cost was then multiplied by optimisation parameters. For *WN-g-sc* and *WN-g-ch* one link direction was preferred. The chaining algorithm set the direction from the newly attached vertex to its fixed predecessor in a polysemy chain. SemCor links were directed also in a reversed order, i.e., from the consecutive sense to the preceding one. These link costs were marked with As. We also inserted the oppositely directed semantic links, marking them with Bs. All links other than polysemy relations (i.e., taxonomic and gloss links) were equipped with the cost of 1. The baseline model *WN-g* achieved in such a setting is $\rho = 0.218$ (see Table 1).

For the complete graph (*WN-g-co*), we applied only one cost parameter A, because the graph edges were bidirectional ($B = A$). The $\rho = \rho(A)$ curve turned to be discontinuous, which is clearly visible in Fig. 2. When the cost of polysemy links was lower than 1 (i.e., the constant cost of WordNet taxonomic and gloss relations), the merged network seemed to perform worse than the *WN-g* baseline. Having passed the threshold of 1, the $\rho(A)$ curve suddenly rose, and reached the maximum value of Spearman's correlation $\rho = 0.226$ for $A \in [1.525,\ 1.650]$, then slowly descended to get to the baseline value 0.218 at the end of inspected area (at $A = 4$). The optimum was obtained in two steps: first, we checked $\rho(A)$ values for $A = 0, 0.25, 0.5, ..., 3.75, 4$, second, we concentrated on the range $A \in [1, 2]$ having thickened the mesh five times ($A = 1, 1.05, 1.1, ..., 2$). Finally, we took the point $A_{opt} = 1.6$ located in the middle of the maximum region as the approximation of the optimal point.

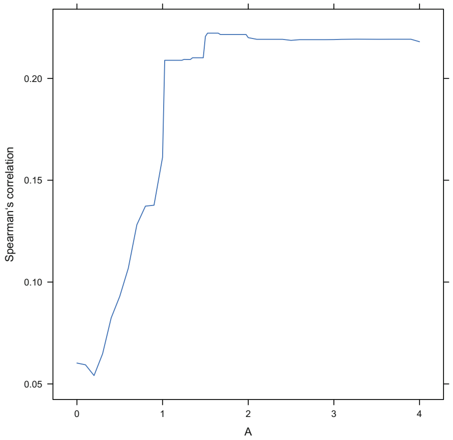

Fig. 2. Spearman's correlation ρ in the function of the cost parameter A for WordNet-gloss graph expanded with complete lemma polysemy nets. Please note that in the case of the *WN-g-co* model $B = A$, due to the impossibility of distinguishing the preferred directions in such a graph.

We tested *WN-g-sc* and *WN-g-ch* models on the mesh of 11 × 11 points, magnifying interesting regions with the denser mesh of 9 × 9 points, checking altogether roughly 200 combinations of parameters A and B. Each net model was optimised with a visual inspection of level plots (Figs. 3 and 4). We tested different values of costs in Dijkstra's algorithm, $\rho = \rho(A, B)$.

For the SemCor-based polysemy graph (*WN-g-sc*), the maximum point was thus identified to be near the point $A = 0.75$, $B = 1$. The chaining-algorithm-based polysemy net (*WN-g-ch*) optimum seemed to be located close to the $(0.425, 0.425)$ point.

Table 1 presents polysemy network sizes together with the estimations of optimal points and maximum Spearman's ρ values. The *WN-g-co* is the biggest one, but the quality of links is not so high ($A = 1.6$, which is greater than the basic cost of 1 for WordNet taxonomic relations and glosses). SemCor-based optimal graph (*WN-g-sc*) received lower cost for A ($A = 0.875$, $B = 1$), which is not surprising since the net, though relatively small, is possibly almost completely error-free. Taking into account that the discovered optimal costs for the *WN-g-ch* graph were the lowest, one might argue that the relation set was the best

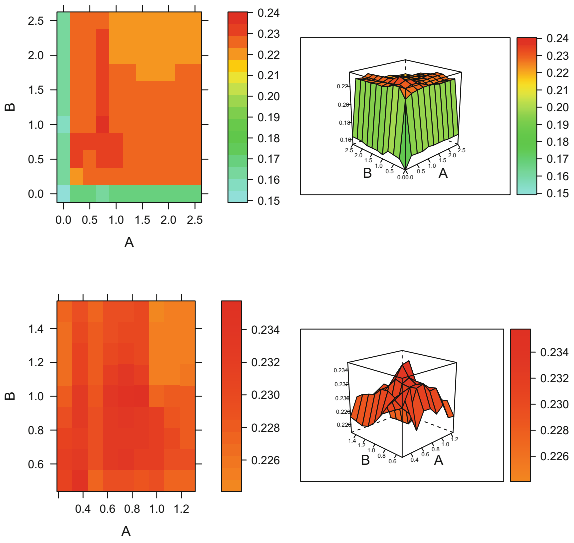

Fig. 3. Spearman's correlation ρ in the function of the cost parameters A and B for WordNet-gloss graph expanded with polysemy links between word senses co-occurring in the very same text in SemCor. Top: Correlations for the square $[0, 2.5] \times [0, 2.5]$. Bottom: 2-times magnification of the optimum area.

model. The graph complements the links for all WordNet senses, while *WN-g-sc* only for those that co-occurred in the corpus.

Choosing the Best Model. We optimised the parameters of the chaining algorithm and two other models. The next step was to evaluate these findings on a larger independent subset of the evocation data set, counting 10,000 evocation pairs, with the optimal parameter settings (Table 1, *evo10k* columns). As a baseline, we have chosen the *WN-g* (the gloss-expanded WordNet graph).

Table 1. Different variants of polysemy nets: *WN-g-co* - the complete graph of inter-sense links, WN-g-sc – the SemCor-based graph of polysemy senses co-occurring in the same text, *WN-g-ch* – the chaining algorithm model with a fixed starting point. Results were obtained on the *evo2k* tuning data set and evaluated on the *evo10k* data set.

Polysemy network	Size $[10^3]$	Vector of costs (w_{WN}, w_g, A, B)	ρ evo2k	ρ evo10k	r evo10k
WN	0.0	(1,1,-,-)	0.138	.149	.215
WN-g	0.0	(1,1,-,-)	0.218	.183	.237
WN-g-co	377.7	(1,1,1.6,1.6)	0.226	.181	.239
WN-g-sc	28.6	(1,1,0.875,1)	0.235	.195	.251
WN-g-ch	110.0	(1,1,0.425,0.425)	0.242	.198	.263

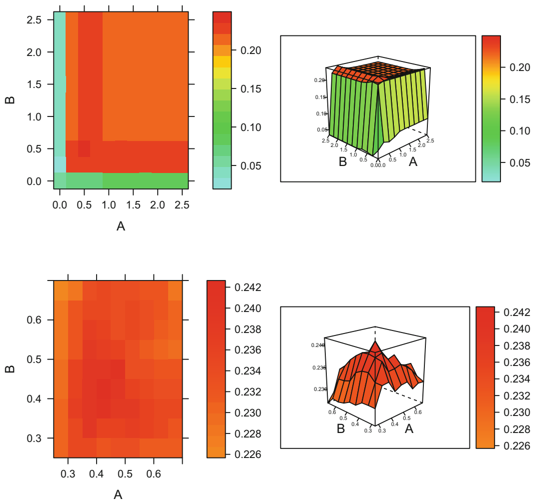

Fig. 4. Spearman's correlation in the function of the cost parameters A and B for WordNet-gloss graph expanded with polysemy links between word senses obtained through the chaining algorithm. Top: Correlations on the square $[0, 2.5] \times [0, 2.5]$. Bottom: 5-times magnification.

The comparison was performed with the use of a bootstrap percentile method ($B = 1,000$ repetitions). The evaluation setting is presented in Table 2 ($m = 7$ is a number of comparisons). Both the chaining algorithm (*WN-g-ch*) and the SemCor-based model (*WN-g-sc*) turned out sufficient in beating the baseline model *WN-g* and a large complete graph (*WN-g-co* model). Polysemy chains were not significantly better than a smaller SemCor co-occurrence model. The semantic network built from SemCor seemed to be incomplete, though. The corpus itself missed many valuable sense associations. Having taken into account the advantages of our chaining procedure which produced at least as good polysemy model and yet a more-complete polysemy network, we decided to choose the *WN-g-ch* model for further experiments.

3.2 Chaining Algorithm in Evocation Recognition

The modified graph *WN-g-ch* contains semantic links introduced by applying the chaining procedure to polysemous words. We use the new resource and our new similarity metric to compute additional feature set for evocation strength prediction. The additional features (mainly the *DSch* feature, but also few frequency-based features) were used as an expansion for the feature set proposed by [11]. We mark the additional features by a cross symbol in Table 3.

The features proposed by Hayashi [11] were based on different language resources. We implemented some of the similarity measures as described in the original work (e.g. the cosine similarity of *AutoExtend* sense representations using the same pre-trained model, symbol: *cosAE*), or replaced the remaining

Table 2. P-values of the one-sided paired bootstrap test for the difference between ρ values (*DSch* measure, $B = 1,000$ repetitions). Alternative hypotheses are formulated in the following manner: 'a row is greater than a column', with an exception of the *WN/WN-g* comparison which has the hypothesis reversed. Statistical significance was calculated through Benjamini-Hochberg procedure at the confidence level of 95% ($m =$ 7 comparisons). We mark statistically significant results with asterisks.

Graph		WN-g	WN-g-co	WN-g-sc
Type	ρ	.183	.181	.195
WN	.149	0.000*	—	—
WN-g-co	.181	0.647	—	—
WN-g-sc	.195	0.014*	0.016*	—
WN-g-ch	.198	0.01*	0.005*	0.354

functions with equivalent measures (e.g., *Wu-Palmer* similarity of lexicalised concepts was replaced by the Jaccard similarity measure, also calculated on WordNet). To predict evocation strength we followed the same approach and we treated the task as a regression problem. We used two different regression models: i) a regressor based on Random Forest framework, and ii) a regressor based on Feed-Forward Neural Network.[6] The final feature set that was used to train our models is presented in Table 3.

- Frequency index $FREQsc(s)$ represents the frequency of a given lemma sense s computed on the basis of the SemCor corpus.
- Frequency-based score $FRANic(s)$ represents a fraction of the overall frequency $FREQic$ of lemma l computed for each of its senses s, where the final score is inversely proportional to sense variant (as they were ordered in WordNet). The lower sense variant number $VAR(s)$ was (e.g. $1^{st}, 2^{nd}, 3^{rd}, ...$), the bigger fraction of the frequency $FREQic(l)$ it received:

$$FRANic(s) = \frac{FREQic(l)}{(VAR(s) + 1)}. \tag{3}$$

  We used the frequencies provided by an internet corpus of English[7].
- The Jaccard Index $JaccSim(s, t)$ is the number of common neighbors divided by the number of nodes that are the neighbors of at least one of input senses (source or target) being considered.
- $Dist(s, t)$ is the length of the shortest path between source and target synset (cf. the description in the previous section).
- $GlossDice(s, t)$ represents Dice similarity measure based on glosses. The measure is computed using all the neighbours in the vicinity of $k = 3$ steps from source and target concepts. We take all of the senses appearing in the glosses

[6] The experimental part was conducted in WEKA framework [8].

[7] Published by Centre for Translation Studies, University of Leeds: http://corpus.leeds.ac.uk/list.html, CC-BY licence.

Table 3. Prediction of evocation strength (individual features): 5-fold cross-validation on the set of 108,000 evocation pairs. Symbols: DV - corpus-based distributional vectors, KB - knowledge-based measures, KV - WordNet-based vector spaces. All numbers represent Pearson's r correlations. Hayashi's data are given for the whole 120,000 pair evocation set. The features marked with an asterisk sign were implemented after [11].

	NN	RF
DV features:		
cosFT	0.1980	0.1247
cosGV	0.2487	0.1547
FRANic	0.0663	0.0626
FREQsc	0.0510	0.0420
KB features:		
Dist	0.1823	0.2476
GlossDice	0.1288	0.0403
JaccSim	0.1131	0.1178
DSch	**0.2596**	**0.2688**
KV measures:		
*cosAE**	0.2122	0.1239
*relVecAE**	0.0341	—
*posSem**	0.1428	0.1731
All features	**0.4415**	0.4363
Hayashi (2016)	**0.4391**	0.3695

of source and target entities as well as the senses from glosses of their neighbours.

- The $posSem(s,t)$ feature is inspired by the work of [11], with a minor alteration – instead of 5 PoS we have 4 PoS.
- $cosFT(s,t)$, $cosGV(s,t)$, and $cosAE(s,t)$ represent the cosine similarity of vector space representations of source and target concepts s and t computed using $fastText$ [2], $GloVe$ [16], and $AutoExtend$ [18] embeddings.
- *relVecAE* - a 300D vector of differences between two AutoExtend vector embeddings (each for one sense in an evocation pair).

The last feature was removed from the RF feature set after preliminary experiments on the *evo10k* data set, since using word (or sense) embeddings directly as feature vector did not improve the quality of the model. In the NN framework the impact of AutoExtend vector differences was positive, though small.

Final Results. In all paired t-tests our *DSch* measure proved to behave better in predicting the evocation strength than any other individual measure at 5% significance level (with Benjamini-Hochberg correction, Table 3). Without Hayashi's original validation folds we were unable to compare the final performance in a direct way – for our NN model we have obtained the absolute mean value of Pearson's r correlation slightly higher than that of Hayashi's NN [11] (see Table 3), but the one-sided one-sample t-test was inconclusive (p-value above 0.05). We were able to prove better performance of both our NN i RF models over Hayashi's RF model at the significance level of 0.001.[8]

4 Conclusions and Further Work

In this paper we presented a novel method of expanding WordNet with polysemy links based on the nearest-neighbor chaining algorithm. We have proven that the new lexical resource facilitates evocation recognition, compared to competitive WordNet-based graphs. We also re-used it successfully within the Neural Network and Random Forest frameworks. We proved that the polysemy-based Dijkstra's distance measure was quite efficient in recognizing evocation, especially when compared to efficiencies of other knowledge-based measures.

Hayashi [11] believed that the real progress in evocation strength recognition could be obtained through merging diverse language resources representing different aspects of human linguistic competence. In this work we focused on a small piece of the puzzle, namely lexical polysemy links. Since we did not utilize *all* Hayashi's features (e.g., the LDA topic modelling measure), there is still space for further improvements.

Applications of the polysemy expanded WordNet go beyond detecting evocation strength. We plan to verify its usefulness in similarity recognition tasks as well as Word Sense Disambiguation.

Acknowledgments. This research was financed by the National Science Centre, Poland, grant number 2018/29/B/HS2/02919, and supported by the Polish Ministry of Education and Science, Project CLARIN-PL.

References

1. Barque, L., Chaumartin, F.R.: Regular polysemy in WordNet. J. Lang. Technol. Comput. Linguist. **24**(2), 5–18 (2009)
2. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606 (2016)
3. Boyd-Graber, J., Fellbaum, C., Osherson, D., Schapire, R.: Adding dense, weighted, connections to WordNet. In: Proceedings of the Global WordNet Conference (2006). docs/jbg-jeju.pdf
4. Cattle, A., Ma, X.: Predicting word association strengths. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1283–1288 (2017)
5. Chklovski, T., Mihalcea, R.: Building a sense tagged corpus with open mind word expert. In: Proceedings of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions, vol. 8, pp. 116–122. Association for Computational Linguistics (2002)

[8] Shapiro-Wilk tests gave p-values equal to 0.4804 (NN) and 0.4923 (RF).

6. Cramer, I.: How well do semantic relatedness measures perform?: A meta-study. In: Proceedings of the 2008 Conference on Semantics in Text Processing, pp. 59–70. Association for Computational Linguistics (2008)
7. Fellbaum, C.: WordNet: An Electronic Lexical Database. The MIT Press, Cambridge (1998)
8. Frank, E., Hall, M., Witten, I.: The WEKA Workbench. Online Appendix for "Data Mining: Practical machine Learning Tools and Techniques". Morgan Kaufmann, Cambridge (2016)
9. Freihat, A.A., Giunchiglia, F., Dutta, B.: A taxonomic classification of WordNet polysemy types. In: Proceedings of the 8th GWC Global WordNet Conference (2016)
10. Geeraerts, D.: Theories of Lexical Semantics. Oxford University Press, New York (2010)
11. Hayashi, Y.: Predicting the evocation relation between lexicalized concepts. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 1657–1668 (2016)
12. Kacmajor, M., Kelleher, J.D.: Capturing and measuring thematic relatedness. Lang. Resour. Eval. **54**(3), 645–682 (2019). https://doi.org/10.1007/s10579-019-09452-w
13. Lipka, L.: An Outline of English Lexicology: Lexical Structure, Word Semantics, and Word-formation, vol. 3. Walter de Gruyter, Berlin (2010)
14. Lyons, J.: Semantics, vol. 2. Cambridge University Press, Cambridge (1977)
15. Nikolova, S.S., Boyd-Graber, J., Fellbaum, C., Cook, P.: Better vocabularies for assistive communication aids: connecting terms using semantic networks and untrained annotators. In: ACM Conference on Computers and Accessibility (2009). docs/evocation-viva.pdf
16. Pennington, J., Socher, R., Manning, C.D.: GloVe: global vectors for word representation. In: EMNLP (2014)
17. Ramiro, C., Srinivasan, M., Malt, B.C., Xu, Y.: Algorithms in the historicalemergence of word senses. Proc. Natl. Acad. Sci. **115**(10), 2323–2328 (2018). https://doi.org/10.1073/pnas.1714730115, https://www.pnas.org/content/115/10/2323
18. Rothe, S., Schütze, H.: Autoextend: extending word embeddings to embeddings for synsets and lexemes. arXiv preprint arXiv:1507.01127 (2015)
19. Youn, H., et al.: On the universal structure of human lexical semantics. Proc. Natl. Acad. Sci. **113**(7), 1766–1771 (2016)