Regular polysemy: from sense vectors to sense patterns

Anastasiya Lopukhina
Neurolinguistics Laboratory,
National Research University
Higher School of Economics (HSE);
Russian Language Institute RAS
alopukhina@hse.ru

Konstantin Lopukhin
Scrapinghub
kostia.lopuhin@gmail.com

Abstract

Regular polysemy was extensively investigated in lexical semantics, but this phenomenon has been very little studied in distributional semantics. We propose a model for regular polysemy detection that is based on sense vectors and allows to work directly with senses in semantic vector space. Our method is able to detect polysemous words that have the same regular sense alternation as in a given example (a word with two automatically induced senses that represent one polysemy pattern, such as ANIMAL / FOOD). The method works equally well for nouns, verbs and adjectives and achieves an average recall of 0.55 and an average precision of 0.59 for ten different polysemy patterns.

1 Introduction

Polysemy is widely spread in natural language. Many studies in linguistics show evidence that certain word classes share polysemy patterns, which means that there are regularities in the way polysemous words vary in their meaning (Shmelyov, 1966; Lakoff and Johnson, 1980; Apresjan, 1995; Pustejovsky, 1995; Paducheva, 1998). These regularities can be explained by analogical processes like semantic shifts (*lamb* can denote either ANIMAL or FOOD), metonymy (*church* can denote either ORGANIZATION or LOCATION) and metaphor (e.g. *dirty* in contexts such as *dirty shoes* and *dirty words*). Because of its significance, regular polysemy has been extensively investigated in lexical semantics (Apresjan, 1971, 1995; Nunberg and Zaenen, 1992; Nunberg, 1995; Uryson, 2003; Zaliznyak, 2006). However, this phenomenon has been little studied in computational semantics and even less in distributional semantics. Several studies that aimed to model regular polysemy in semantic vector space were focused on word vectors. Del Tredici and Bel (2015) proposed a method, based on word embeddings and regular semantic alternations, that allows detecting polysemous nouns among all nouns and representing them in a way that accounts for asymmetry in sense predominance. Di Pietro (2013) detected sense alternations by performing word sense disambiguation using vectors of words that denoted sense domains, such as ANIMAL or FOOD. Boleda and colleagues (2012b) compare word vectors for polysemous nouns with average vectors of monosemous words in predefined sense domains. Their study relied on the CoreLex meta sense inventory that was built using WordNet. Thus, the aforementioned methods all use semantic word vectors to detect sense alternations.

For the task of regular polysemy detection, we use sense vectors and not word vectors. We believe that this is a more natural approach to the problem, because it allows us to study regular sense alternations as they are: we deal directly with senses and their location in semantic vector space. Our approach has two major advantages: first, we believe that it is less affected by sense skewness than methods based on word vectors, because vectors of different senses are distinct, even if senses have very different frequencies, while in case of word vectors, a much more frequent sense will determine the word vector, as noted by Del Tredici and Bel (2015). Second, theoretically our approach is not limited by regular alternation between just two senses, as in previous studies (Boleda et al., 2012a, 2012b; Vieu et al., 2015), but can be naturally extended to three or more senses.

The sense vectors we use in our study are built automatically on a big corpus. The technique of automatic word sense induction (WSI) allows us to represent senses as clusters of semantically similar instances. Usually, the technique does not use any external resources such as dictionaries, thesauri or sense-tagged data. WSI was successfully applied to the lexicographical task of novel sense detection, i.e. identifying words which have taken on new senses over time (Lau et al., 2012). Besides, WSI provides data for the study of diachronic variation in word senses (Bamman and Crane, 2011). Although Boleda and colleagues (2012b, p. 153) noted that automatic word sense induction could lead to more flexible and realistic models of regular polysemy, to the best of our knowledge, the WSI technique was not used in any previous research of this type.

In this study, we propose a model for regular polysemy detection that is based on sense vectors and allows us to work directly with senses in semantic vector space. We performed an experiment on Russian nouns, verbs and adjectives and subsequently discuss the limitations of our method.

2 Method

The core of our method can be described as follows: in semantic vector space we take two senses of a word (s_a and s_b; with predefined regular alternations such as ANIMAL / FOOD) and search for similar pairs (s_i, s_j) of another word, where s_a is close to s_i and s_b is close to s_j. This approach is very similar to how regular polysemy is defined in (Apresjan, 1995, p. 189). To be more precise, given word w and its senses s_a and s_b, and another candidate pair of senses (s_i, s_j) for word w_k, we define their pattern similarity measure $PatternSim$ as:

$$PatternSim((w, s_a, s_b), (w_k, s_i, s_j)) = \min(sim((w, s_a), (w_k, s_i)), sim((w, s_b), (w_k, s_j))) \quad (1)$$

where sim is a cosine similarity measure between sense vectors, and (w, s) is a sense vector. Using this similarity measure between pairs of senses, we take the top N_{lim} of all possible candidate pairs having similarity above threshold δ, where N_{lim} and δ are hyperparameters of the method.

Sense vectors are produced by the adaptive Skip-gram model AdaGram, which is a non-parametric extension of Skip-gram word2vec model to word senses. It automatically learns the required number of representations for all words at a desired semantic resolution (Bartunov et al., 2015). It is able to learn a dense vector embedding for each sense of a word, where the number of senses is determined using a constructive definition of the Dirichlet process via the stick-breaking representation. AdaGram has an efficient online learning algorithm and was evaluated on word sense induction tasks of SemEval-2007 and 2010 (Bartunov et al., 2015, pp. 8-9). Hyperparameter α controls granularity of produced senses, and other hyperparameters, such as vector dimension and window size, have the same role as in word2vec algorithm. Sense vectors produced by AdaGram can be represented as words that are nearest neighbors (e.g. *Monty*, *Perl*, *Molurus* for different senses of the word *python*) or by context words with the highest PMI.

In (Lopukhina and Lopukhin, 2016) a qualitative and quantitative evaluation of several WSI methods, including AdaGram, was performed on 15 Russian nouns. Other methods included LDA, context clustering, clustering of word2vec neighbors. For the quantitative evaluation, the authors measured similarity of suggested clustering to the existing dictionary senses with Adjusted Rand Index (ARI) and V-measure scores. For the qualitative evaluation, they assessed the interpretability of the derived senses, the number of duplicate senses, the number of mixed senses and derivation of rare senses. Trained on a 2 billion word Russian corpus with $\alpha = 0.15$, AdaGram discovered the largest number of senses, and was a close second in both ARI and V-measure. Compared to context clustering, which was first in the quantitative evaluation, AdaGram is much more efficient and was able to discover even rare senses.

3 Experiment

We aimed to study how well the proposed technique detects polysemous words that have the same regular sense alternation as in a given example. An example is a word with two automatically induced senses that represent one polysemy pattern (such as ANIMAL / FOOD). We manually selected ten polysemy patterns: four for nouns, three for verbs and three for adjectives, nine of them from the most famous

and reliable description of regular polysemy for Russian, *Lexical Semantics* by Jury Apresjan (1995), in which he thoroughly classifies and illustrates more than 80 productive and non-productive regular polysemy types for the aforementioned parts of speech. We also took one polysemy pattern for verbs from an ongoing work led by Valentina Apresjan (2016). Besides, we checked that word senses, which were part of the polysemy pattern in question, were presented in the corpus and were thus detected by the AdaGram model (e.g. *zheleznyj* 'iron': *iron gates* is sense #3 in AdaGram / *iron will* is sense #4 in AdaGram).

Polysemy patterns for nouns:
ANIMAL / FOOD (e.g. *gus'* 'goose');
AMOUNT / CONTAINER (e.g. *butylka* 'bottle');
ACTION / RESULT (e.g. *ushyb* 'injury' / 'bruise');
MUSIC / DANCE (e.g. *val's* 'waltz ').

Polysemy patterns for verbs:
AUTONOMOUS RELOCATION / NONAUTONOMOUS RELOCATION (e.g. *jehat'* 'to move (about a car)' / 'to drive (a car)');
PRODUCE SOUND / SPEAK (e.g. *blejat'* 'to bleat');
CEASE TO EXIST / RUN OUT OF INNER RESOURCE (e.g. *tajat'* 'to melt' / 'to melt away').

Polysemy patterns for adjectives:
MADE OF SOME MATERIAL / MAKING A SIMILAR IMPRESSION (e.g. *derevjannyj* 'wooden');
SURFACE PROPERTY / HUMAN PROPERTY (e.g. *nezhnyi* 'delicate');
HAVING SOME TASTE / MAKING A SIMILAR IMPRESSION (e.g. *kislyj* 'sour').

Then, for each pattern we selected 4-7 examples that were used for evaluation. All the examples were extracted from *Lexical Semantics* (1995) or from (Apresjan, 2016). We were guided by the following principle: Words should be semantically similar, namely synonyms, antonyms or co-hyponyms. In the study by Jury Apresjan (1995), polysemy patterns such as ACTION / RESULT embrace a large number of semantically very different words from *ushyb* 'injury' / 'bruise' to *risunok* 'drawing' and *ispravleniye* 'correction'. For the purpose of the present study, we chose words from one semantic domain (e.g. *ushyb* 'injury' / 'bruise'; *ukus* 'bite' / 'wound'; *perelom* 'breaking of a bone' / 'fracture'; *porez* 'cut', words denoting different injuries and their result on/in the human body).

The experimental setup was as follows: Sense vectors were built using AdaGram with $\alpha = 0.10$, window size 5, vector dimension 300, maximum number of senses 10 and minimal token frequency 100. Corpus used for training contained about 2 billion tokens and was a combination of ruWac (a representative snapshot of the Russian Web), lib.ru (a Russian online library) and Russian Wikipedia. Corpus was lemmatized with Mystem 3, lowercased and cleared of punctuation.

4 Evaluation

In order to study how well our method is able to detect word sense alternations, we evaluated recall and precision in two separate experiments. In both cases, two words were selected from each polysemous pattern group as "anchor" words, while other words of the group were treated as "target" words. Each of the anchor words (with its two senses) was given as input to the method, thus defining a sense alternation by an example.

In the recall evaluation we checked how many of the target words were actually produced, given the anchor word. Recall was evaluated with two different limits on the number of detected words N_{lim}, 5 and 50. Note that we did not expect a high recall with $N_{lim} = 5$, as we believe that there can be other words besides target words that have the same alternation. Another reason is that there were sometimes more than 5 target words in the group, therefore it was impossible to achieve perfect recall with just $N_{lim} = 5$. Different parts of speech did not show significant variation of recall. The average recall for ten groups was **0.22** for $N_{lim} = 5$ and **0.55** for $N_{lim} = 50$.

In order to evaluate precision, we took anchor words and for each of them extracted the top five candidates ($N_{lim} = 5$) that were produced by our method. These candidates were checked by a lexicographer:

if a candidate shared the same polysemy pattern with the anchor word, it was accepted. The average precision for ten groups was **0.59**, with three groups having perfect precision. In most cases, words that were rejected were semantically similar to the anchor word, but they did not exhibit the polysemy pattern in focus. For example, wrong candidates for the word *ukus* 'bite' / 'wound' were *snake*, *insect* and *mosquito*, which can be subjects of the action; wrong candidates for the word *stakan* 'glass' were *tea* and *coffee*, which denote the content; and wrong candidates for the word *kislyj* 'sour' were *garlicky* and *fried*, which mean HAVING SOME TASTE, but do not exhibit the meaning MAKING SIMILAR IMPRESSION.

5 Discussion

The method for detecting words of a predefined polysemy pattern showed promising results in both precision and recall. The method allows obtaining words with the same sense alternations, given one example directly from the corpus, and works equally well for nouns, verbs and adjectives. However, the method we propose has limitations that can be explained by the nature of the method and the way distributional models are built.

One of these limitations is that some senses of words that are a part of a polysemy pattern can hardly be distinguished by means of the distributional model and are not clearly represented in a vector space model. For example, many verbs, as described in *Lexical Semantics* (1995), have a 'causation' component in one of their meanings, e. g. *lit'* 'to pour' in contexts such as *He poured the last of the water down the sink* and *The water pours from the tap*. These two senses can be distinguished syntactically or by taking word order into account, but this cannot be achieved by our proposed model. Some verbs differ in the properties of the objects they attach, e. g. *varit'* 'to cook' in contexts like *to cook potatoes* (potato changes its properties) and *to cook soup* (soup appears); this difference cannot be captured by our model. The problem of sense discrimination by context is most evident for verbs.

Another limitation is caused by the difference between the notion of "regular polysemy" in theoretical studies and in its computational implementation. Lexicologists formulate sense alternation principles in a very general sense and thus, semantically different words may have the same polysemy pattern, e.g. adjectives *gornyj* 'alpine' in contexts such as *alpine range* and *alpine ski*, *glaznoj* 'ocular' / 'eye' in contexts such as *ocular fundus* and *eye drops*, and *krysinyj* 'rat' in contexts such as *rat tail* and *rat poison* share the same pattern RELATING TO SMTH / DESIGNED FOR SMTH. Semantically different words cannot be detected with distributional models because they appear in different contexts, which means that the method we propose is limited by synonyms, antonyms and co-hyponyms.

We believe that our model for regular polysemy can also be applied to an unsupervised discovery of patterns. The $PatternSim$ measure defined above (eq. 1) can be used to cluster all pairs of a particular part of speech, hence each cluster will represent a distinct regular polysemy pattern. Another possible extension is to change $PatternSim$ in a way that will account for the direction of a vector between two senses. Given two senses s_a and s_b of word w and another pair of senses s_i and s_j of word w_k, we believe that these two pairs of senses will be more similar if vectors $s_a - s_i$ and $s_b - s_j$ have similar directions.

6 Conclusion

In this study, we describe an approach to the automatic detection of regular sense alternations from the corpus given an example. Our approach is based on sense vectors and gives the opportunity to deal with senses directly. It allows finding semantically similar words that share the same polysemy patterns. It works equally well for nouns, verbs and adjectives and achieves an average recall of 0.55 and average precision of 0.59 for ten different polysemy patterns.

Our model uses sense vectors that are produced with the AdaGram method and, being a distributional model, does not fully cover all types of regular alternations that are described in the theoretical literature; it is only applicable to sense alternations in semantically similar words.

The method we describe can be useful for theoretical studies of regular polysemy and for lexicographers. It is available online at `http://adagram.ll-cl.org/about`.

Acknowledgements

References

Apresjan, Jury D. 1971. Regular polysemy. *Proceedings of the Academy of Sciences of the USSR. Department of Literature and Language.* Vol. 30, Moscow. pp. 509-523.

Apresjan, Jury D. 1995. *Lexical Semantics. Selected works.* Volume I, Moscow.

Apresjan, Valentina Ju, 2016. Ischeznut' 'to disappear' and propast' 'to vanish': polysemy and semantic motivation. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2016"*, pp. 16-28, Moscow.

Bamman, David and Gregory Crane. 2011. Measuring historical word sense variation. *Proceedings of the 2011 Joint International Conference on Digital Libraries (JCDL 2011)*, pp. 1-10, Ottawa, Canada.

Bartunov, Sergey, Dmitry Kondrashkin, Anton Osokin, and Dmitry Vetrov. 2015. Breaking sticks and ambiguities with adaptive skip-gram. Accessed September 17, 2016. `https://arxiv.org/abs/1502.07257`.

Boleda, Gemma, Sabine Schulte im Walde, and Toni Badia. 2012a. Modeling regular polysemy: A study of the semantic classification of Catalan adjectives. *Computational Linguistics* 38:3, pp. 575-616.

Boleda, Gemma, Sebastian Padó, and Jason Utt. 2012b. Regular polysemy: a distributional model. *First Joint Conference on Lexical and Computational Semantics* (*SEM), pp. 151-160, Montral, Canada.

Del Tredici, Marco and Núria Bel. 2015. A Word-Embedding-based Sense Index for Regular Polysemy Representation. *Proceedings of NAACL-HLT*, pp. 70-78.

Di Pietro, Giulia. 2013. *Regular polysemy: A distributional semantic approach.* (2013). Master thesis, Universit`a di Pisa.

Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By.* University of Chicago Press.

Lau, Jey Han, Paul Cook, Diana McCarthy, David Newman, and Timothy Baldwin. 2012. Word sense induction for novel sense detection. *In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 591-601. Association for Computational Linguistics.

Lopukhina, Anastasiya and Konstantin Lopukhin. 2016. Word sense induction methods: which one is better for Russian. Accessed October 1, 2016. `https://www.academia.edu/26080939/Word_Sense_Induction_Methods_Which_One_Is_Better_for_Russian`.

Nunberg, Geoff and Annie Zaenen. 1992. Systematic polysemy in lexicology and lexicography. *Proceedings of Euralex II*, pp. 387-395, Tampere, Finland.

Nunberg, Geoff. 1995. Transfers of meaning. *Journal of Semantics*, 12(2), pp. 109-132.

Paducheva, Elena V. 1998. Paradigms of semantic derivation for Russian verbs of sound. *Proceedings of Euralex VIII*, v. 1, pp. 231-238, Liège, Belgium.

Pustejovsky, James. 1995. *The Generative Lexicon.* The MIT Press, Cambridge, MA.

Shmelyov, Dmitrij N. 1966. On the analysis of word semantic structure. *Zeichen und System der Sprache.* Bd. 3. Berlin.

Uryson, Elena V. 2003. *Problems in linguistic worldview studies: Analogy in semantics.* Languages of Slavic Cultures, Moscow.

Vieu, Laure, Elisabetta Jezek, and Tim Van de Cruys. 2015. Quantitative methods for identifying systematic polysemy classes. *Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics.* Tubingen.

Zaliznyak, Anna A. 2006. *Ambiguity in language and methods for its presentation.* Languages of Slavic Cultures, Moscow.