

# Visualising WordNet Embeddings: some preliminary results

Csaba Veres

Department of Information Science and Media Studies

The University of Bergen, Norway

csaba.veres@uib.no

## Abstract

AutoExtend is a method for learning unambiguous vector embeddings for word senses. We visualise these word embeddings with t-SNE, which further compresses the vectors to the  $x,y$  plane. We show that the t-SNE co-ordinates can be used to reveal interesting semantic relations between word senses, and propose a new method that uses the simple  $x,y$  co-ordinates to compute semantic similarity. This can be used to propose new links and alterations to existing ones in WordNet. We plan to add this approach to the existing toolbox of methods in an attempt to understand learned semantic relations in word embeddings.

## 1 Introduction

There is currently a great deal of interest in the representations of words as continuous, real valued vectors, or *embeddings*. Various popular methods produce a single vector for each word form in the training set, for example GloVe (Pennington et al., 2014), word2vec (Mikolov et al., 2013a), and SVD (Levy et al., 2015).

These methods could be regarded as modern day experiments inspired by Zellig Harris' hypotheses about the distributional structure of language. Harris proposed that word meanings give rise to observable distributional patterns in language, such that two semantically unrelated words A and C would be less likely to be found in common linguistic contexts as two semantically related words A and B (Harris, 1954). Modern machine learning techniques have made it computationally possible to *embed* very high dimensional distributional patterns in a much lower dimensional vector space, in which the distances between any given vectors is related to the similarities of context in which the corresponding words

are found in the training set. Semantic relatedness is therefore correlated with the calculated distance (e.g. cosine distance) between vectors, although the precise nature of the relatedness is not well understood. One of the long term motivations behind the work reported in this paper is to develop a methodology for investigating the nature of the semantic relationships discovered by various methods of context embedding.

A general problem with current methods of single layer *embeddings* is that they treat each word-form as a single word in a *bag of words* model. Thus the embedding for each word-form conflates contexts over every sense of ambiguous words. There have been proposals to discover unique vectors for the different senses of ambiguous words, typically by using clusters of words related to the different senses, either before (Reisinger and Mooney, 2010) or after training (Schütze, 1998).

In this paper we investigate semantic relationships between WordNet synsets using word embeddings. The most convenient resource for this are the vectors trained with AutoExtend (Rothe and Schütze, 2015). This method uses structural information from WordNet to learn new embeddings for synsets and lexemes from non-disambiguated word vectors. Their insight is to use the constraints detailed in WordNet<sup>1</sup>, and to formalise those constraints with respect to the embeddings. For example, the learned embedding for the word-form *W/suit* is formally related to two lexemes, one *L/suit (S/suit-of-clothes)*, and the other *L/suit (S/lawsuit)*, where the *S* prefix denotes that the lexeme is a part of the synset *S/*. Further, the embedding for the lexeme *L/suit (S/lawsuit)* is connected to the embeddings for the lexemes *L/case (S/lawsuit)* and *L/lawsuit (S/lawsuit)* because they are elements in the synset *S/lawsuit*.

---

<sup>1</sup>The technique is not restricted to WordNet, but could be used with any other resource that defines structural constraints between senses.

Finally, these lexemes are themselves aligned with the words *W/case* and *W/lawsuit*, for which embeddings have been learned (see (Rothe and Schütze, 2015), figure 1). The goal is to learn embeddings for the lexemes and synsets from the embeddings of the words and the formal constraints taken from the resource, in this case WordNet.

The main goal in this paper is to explore semantic relationships in the vector space of lexemes created by the disambiguation algorithm. We compare these to the baseline embeddings created with the word2vec skip-gram model (Mikolov et al., 2013b). To the best of our knowledge the semantics of vector similarities in embedding space have not been subject to rigorous linguistic investigation. We think that investigating semantic relations using the lexemes learned through the AutoExtend framework will provide important data for understanding the relations captured by word embedding techniques in general. We begin with some visualisations before moving on to some more quantitative accounts. The experiments reported in this paper are at an early stage, mostly aimed at gathering observations rather than finding explanations for them.

## 2 Lexeme Visualizations

In these experiments we used AutoExtend to learn vectors for 73747 lexemes from embeddings generated with the GoogleNewsCorpus, and WordNet3.0. The first experiment was to visualize the whole set with the T-distributed Stochastic Neighbor Embedding (t-SNE) method (van der Maaten and Hinton, 2008), which is a nonlinear dimensionality reduction technique that attempts to keep the relative distances in the high dimensional space intact during the low dimensional transformation. Perhaps not surprisingly the visualisation of the entire set was not terribly useful because of its very high density of points, and is not reproduced here.

The second experiment was to visualize a meaningful sub set of the embeddings that illustrate a sub domain of interest<sup>2</sup>. We took the meaningful subset from an experimental semantic bookmarking platform, LexiTags (Veres, 2013; Veres, 2011), in which users assign WordNet lexemes as *tags* to their bookmarks. The tags are meaningful because

---

<sup>2</sup>a common approach in practice according to an article by Sergei Smetanin in Medium: Towards Data Science <https://is.gd/UyUrhP>

they are used to describe web resources of interest to users of the platform. We collected 248 tags and constructed a t-SNE plot of the corresponding WordNet embeddings (figure 2).

Some interesting relations are immediately apparent. For example the tag *boring* is used in an uncommon sense denoted by the lexeme {boring.n.02: (the act of drilling a hole in the earth in the hope of producing petroleum)}, which in the visualization is closely related to {extraction.n.03: the action of taking out something (especially using effort or force)}. However in the baseline word2vec embeddings only the more common adjectival sense is available, with the related words being {uninteresting, depressing, and dull}.

There also appears to be a cluster that captures an interesting progression from {crime.n.01: an act punishable by law} to {corruptness.n.01: the state of being corrupt}, {government.n.01: the organization that is the governing authority of a political unit} and finally to the result, a {revolution.n.02: the overthrow of a government by those who are governed}. Perhaps a sense of causality between the lexemes has been captured.

Additionally there are some interesting relationships between lexemes from different word classes, for example the actions {synchronize.v.01: make synchronous and adjust in time or manner}, and {install.v.01: set up for use} when used in the domain of computer science often involves in the creation of a {backup.n.04: a copy of a file or directory on a separate storage device}. Again this might be an act of causation.

## 3 Sense Clusters

The visualisations suggest some interesting patterns in the relationships. However a more systematic study will require better ways to quantify observations. To this end we propose a unique method for using the t-SNE results which, to our knowledge, has never been reported.

Recall that the t-SNE algorithm compresses the 300 dimensional vectors into two points  $(x_1, y_1), (x_2, y_2)$  for visualisation, where the distance  $d = |\mathbf{x} - \mathbf{y}|$  is optimised to preserve the neighbourhood relations in the original high dimensional vector space. Thus the distance  $d$  is construed as the semantic distance between the two points. We propose to use these distances directly in calculating the semantic similarities between lexemes, to take the place of cosine similar-



ity in the original vector space. Thus, we have two measures of similarity, which might reveal different clusters.

In order to discover clusters in the  $x,y$  coordinate space we used the divide and conquer approach to the closest pair of points problem, where the closest pair is recursively identified by finding the closest pair in one half of the gradually diminishing problem space<sup>3</sup>. We used a python implementation of the algorithm<sup>4</sup> to find the closest pair of points, then found the five closest points to the first in the pair. Then we deleted one of the closest match points from the initial pair and repeated the divide and conquer algorithm to find the next closest pair of points from the remaining set. In the end this gave us a large set of clusters formed by the closest points in the entire co-ordinate system, and the five closest points to those.

Table 1 shows some hand selected examples of the closest points, together with their neighbours in the two dimensional t-SNE space, the original 300 dimensional AutoExtend space, as well as the word2vec embedding.

It is clear that both sets of results based on the AutoExtend vectors are better able to capture the precise meaning of the search terms, and return more relevant neighbours than word2vec. Common embedding techniques such as word2vec can return words in the result set that are either irrelevant, relevant along some obscure semantic dimension, or simply morphological derivatives of the search term. There are examples of each of these in our result set.

Looking at the two result sets from the lexeme embeddings it appears that the t-SNE results are superior, at least for these examples, to cosine similarity measures. More of the results seem to capture the precise meaning of the particular lexeme. For the *opposite* example, the t-SNE results better capture the sense that *opposites* are *different*. AutoExtend also captures this but to a lesser extent, where the closest neighbour is *identical*, which is the opposite of *opposite*. Right semantics, wrong polarity.

Another interesting observation is that the t-SNE results might be useful in identifying synsets with very similar meanings in WordNet, which is necessary for creating new versions with less fine-

<sup>3</sup>[https://en.wikipedia.org/wiki/Closest\\_pair\\_of\\_points\\_problem](https://en.wikipedia.org/wiki/Closest_pair_of_points_problem)

<sup>4</sup>[http://www.rosettacode.org/wiki/Closest-pair\\_problem](http://www.rosettacode.org/wiki/Closest-pair_problem)

grained meaning distinctions (e.g. (Snow et al., 2007)). Again in the *opposite* example the second and fifth meaning of *different* appear as if they could be merged. The rule would be to merge the synsets for lexemes of the same word form in a cluster.

The next steps in this research is to quantify the relationship between the lexemes in the t-SNE clusters and existing WordNet links. It seems clear from the examples that the embedding relations are not identical to the relations already in WordNet, but can potentially reveal interesting, additional thematic links. This can be used to propose new links in WordNet.

## 4 Conclusion

In conclusion this very brief look at the results shows that the t-SNE compression provides a very interesting set of results to complement the study of semantic relations. As far as we know these are novel ideas which have not been investigated.

We plan to use these results to modify WordNet by merging similar synsets, and by including new thematic links.

Clearly the work is at an early stage, but we are excited at the possibilities presented by these preliminary results.

## References

- Zellig S. Harris. 1954. Distributional structure. *WORD*, 10(2-3):146–162.
- Omer Levy, Yoav Goldberg, and Ido Dagan. 2015. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3:211–225.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Neural and Information Processing System (NIPS)*.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- J Reisinger and RJ Mooney. 2010. Multi-prototype vector-space models of word meaning.

- S Rothe and H Schütze. 2015. Autoextend: Extending word embeddings to embeddings for synsets and lexemes.
- Hinrich Schütze. 1998. Automatic word sense discrimination. *Computational linguistics*, 24:97–123.
- Rion Snow, Sushant Prakash, Daniel Jurafsky, and Andrew Y. Ng. 2007. Learning to merge word senses. In *EMNLP-CoNLL*.
- Laurens van derMaaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *JournalofMachine-LearningResearch*, 9:2579–2605.
- Csaba Veres. 2011. Lexitags: An interlingua for the social semantic web. In *Proceedings of the 4th International Workshop on Social Data on the Web, SDoW@ISWC 2011, Bonn, Germany, October 23, 2011*.
- Csaba Veres. 2013. Crowdsourced semantics with semantic tagging: "don't just tag it, lexitag it!". In *Proceedings of the 1st International Workshop on Crowdsourcing the Semantic Web, Sydney, Australia, October 19, 2013*, pages 1–15.

Table 1: Selected lexemes and their closest neighbours in the t-SNE compression. Also shown are the nearest neighbours in the original AutoExtend embeddings, and the closest neighbours in word2vec. The first row is the target word, neighbours ordered by descending similarity.

| <i>t-SNE most similar words</i>   | <i>most similar words in AutoExtend vector space</i>   | <i>word2vec most similar words</i> |
|---|--|------------------------------------|
| <b>opposite.s.04</b> the other one of a complementary pair; "the opposite sex"  |  |                                    |
| <b>opposite.s.03</b> moving or facing away from each other; "looking in opposite directions" the act of linking things together | <b>opposite.s.01</b> being directly across from each other; facing                               | perpendicular                      |
| <b>different.s.02</b> distinctly separate from the first; "that's another (or different) issue altogether"                      | <b>identical.s.02</b> being the exact same one; not any other                                    | side                               |
| <b>different.s.05</b> distinct or separate; "each interviewed different members of the community"                               | <b>vocationally.r.01</b> affecting the pursuit of a vocation or occupation                       | inwards                            |
| <b>face-to-face.r.02</b> directly facing each other   | <b>variant.s.01</b> differing from a norm or standard  | diagonally                         |
| <b>other.a.01</b> the act of tying or binding things together   | <b>different.s.02</b> distinctly separate from the first   | right                              |
| <b>listening.n.01</b> the act of hearing attentively  |  |                                    |
| <b>sensing.n.02</b> becoming aware of something via the senses  | <b>percussion.n.04</b> tapping a part of the body for diagnostic purposes                        | listened                           |
| <b>taste.n.07</b> a kind of sensing distinguishing substances by means of the taste buds  | <b>auscultation.n.01</b> listening to sounds within the body (usually with a stethoscope)        | listens                            |
| <b>lipreading.n.01</b> perceiving what a person is saying by observing the movements of the lips                                | <b>moralism.n.02</b> judgments about another person's morality                                   | listener                           |
| <b>fingering.n.02</b> touching something with the fingers   | <b>lipreading.n.01</b> perceiving what a person is saying by observing the movements of the lips | hear                               |
| <b>swell.n.03</b> a crescendo followed by a decrescendo   | <b>rehearing.n.01</b> the act of hearing again   | vocalizing                         |