Distributed language models for psycholinguistics analysis

Márton Makrai
Research Institute for Linguistics, Budapest

June 26, 2015

Vector space language models (VSM) constitute a tool for the research of how the meaning of linguistic units (mainly word forms) is represented in the cognitive faculty. VSMs represent words by vectors (more precisely, dense real-valued vectors of some hundred dimensions) trained on corpora with the methods of machine learning. These vectors can be compared to vectors recorded in brain imaging experiments. We report a pilot experiment in this vein, and propose more complex variants in the fields of morphology and word ambiguity, where it can be tested through VSMs how homomorphic the relationship between the structure discovered by theoretical linguists and that implemented in the brain is.

VSMs are useful *in natural language processing* because they map similar words to similar vectors, where the similarity of vectors is measured by their cosine. The mapping is extended to greater units, web queries, short sentences or even paragraphs in successful applications. Compositionality works even within the word: more recent VSMs (Mikolov et al., 2013c) show *relational similarities* (the term is from Turney (2006)) such that the difference between *queen* and *king* is similar to that between *woman* and *man*, or the difference between *years* and *year* is similar to that between *tables* and *table*:

$$\mathbf{v}\,(\text{queen}) - \mathbf{v}\,(\text{king}) \approx \mathbf{v}\,(\text{woman}) - \mathbf{v}\,(\text{man})$$
$$\mathbf{v}\,(\text{years}) - \mathbf{v}\,(\text{year}) \approx \mathbf{v}\,(\text{tables}) - \mathbf{v}\,(\text{table})$$

These relationships enable us to test the compositionality of word formation e.g. affixation, derivation, compounding, or formation of multi word expressions.

Psycholinguistic reality of VSMs has been tested in experiments where human judgements about synonymy, category membership or attribuation (whether ‘father’ is ‘strong’) were compared to those computed from VSMs. A more direct insight of the cognitive process is offered by neural imaging data that also offers itself for comparison with VSM representations of words.

Connecting VSMs to brain imaging data has its practical motivation outside psycholinguistics as well: in theoretical linguistic terms, the lexical meaning of words (which, in the structuralist tradition, is totally determined by their relationship as they appear in texts) has to be *grounded* in the real word. A great part of our knowledge about the meaning of words (e.g. the difference between a dolphin and a whale) resides in the *visual* word. Another motivation comes from web search, namely when users search for images using text. Another application, yet a music of the future, is question answering from pictures like *Is the cat on the table?*.

Besides reporting pilot studies in training a mapping between VSM vectors and brain imaging vectors applying the method of Mitchell et al. (2008) to data trained with models developed in the past years (Mikolov et al., 2013a; Pennington et al., 2014), we propose experiments in morphology and ambiguity. *Morphology* is an eternal topic in distributed language modeling going back to the (in)famous past tense debate with two early representatives of the parties being Pinker (1984) and Rumelhart and McClelland (1986). The first experiment about the productivity of affixation in present-day scale is Lazaridou et al. (2013).

Our third topic is *word ambiguity*. Both traditional lexicons and computational lexical resources like WordNet (Miller, 1995) list several uses of word forms, while the creativity with which speakers use words requires a more abstract treatment of word meaning in computational understanding of human language. The question is made more difficult by the problem of multilinguality: if we are seeking for a kind of word meaning that is, to some degree, independent of the specific language, we find that e.g. a *window* in the outer wall of a building and in a ticket office are conceptually similar but expressed with other words in German *Fenster, Schalter*. Huang et al. (2012) train a VSM with possibly more vectors for the same word form, offering an objective means for drawing the border between polisemy (more uses of the same concept, one vector) and homonymy (more concepts with the same form, more vectors). Brain imaging studies have separated phases of word comprehension before and after disambiguation. This is an other point where VSM vectors and brain imaging data can be compared.

For operationalizing the research in ambiguity, we propose to compare two sets of data. One is the vectors of Huang et al. (2012) who face ambiguity by training possibly more vectors for the same English word form. Based on this data, we can call a word form ambiguous if they have learned more vectors for it, and the vectors differ significantly. For the data to which these disambiguated vectors are to be compared, we propose two choices: (1) We call a word like *window* ambiguous if, in a word translation experiment like Mikolov et al. (2013b) to a VSM for some other language, the vectors for its different uses map to vectors of different words in the target language. (2) Psycholinguistic experiments have been conducted to record brain activations before and after a supposed word disambiguation phase of sentence comprehension. Disambiguated vectors can be compared to data from after the disambiguation phase.

References

Eric H. Huang, Richard Socher, Christopher D. Manning, and Andrew Y. Ng. Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, ACL '12, pages 873–882, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics. URL http://dl.acm.org/citation.cfm?id=2390524.2390645.

Angeliki Lazaridou, Marco Marelli, Roberto Zamparelli, and Marco Baroni. Compositional-ly derived representations of morphologically complex words in distributional semantics. In *ACL (1)*, pages 1517–1526, 2013.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In Y. Bengio and Y. LeCun, editors, *Proc. ICLR 2013*, 2013a.

Tomas Mikolov, Quoc V Le, and Ilya Sutskeve. Exploiting similarities among languages for machine translation. Xiv preprint arXiv:1309.4168, 2013b.

Tomas Mikolov, Wen-tau Yih, and Zweig Geoffrey. Linguistic regularities in continuous space word representations. In *Proceedings of NAACL-HLT-2013*, pages 746–751, 2013c.

George A. Miller. Wordnet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

T. M. Mitchell, S.V. Shinkareva, A. Carlson, K.M. Chang, V.L. Malave, R.A. Mason, and M.A. Just. Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880):1191, 2008.

Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, 2014.

Steven Pinker. *Language Learnability and Language Development*. Harvard University Press, 1984.

D. Rumelhart and J. McClelland. On learning the past tenses of English verbs. In D. Rumelhart and

J. McClelland, editors, *Parallel distributed processing: Explorations in the microstructure of cognition*. Branford, Cambridge MA, 1986.

Peter D. Turney. Similarity of semantic relations. *Computational Linguistics*, 32:379–416, 2006.