Making History Not Count: Should Historical Corpora Really Be Treated Differently for Event Detection Tasks?

Andrea Ferretti

University of Milan, Milan, Italy
andrea.ferretti1@studenti.unimi.it

Abstract. The abstract should briefly summarize the contents of the paper in 150–250 words.

Keywords: Event Detection · Historical Event Detection · Glove · Embeddings comparison.

1 Introduction

Word embeddings aim to map words of a vocabulary to vectors of numbers. This is done in order to have a more versatile, tractable, and mathematical representation of those words and improve the performances of several natural language processing tasks. The more intuitive way to do so would be to have a one-hot encoding representation of each word in the vocabulary. This method, however doesn't provide any information about the meaning of a word. To allow the vectors to retain semantic meaning the foundamental idea of distributional semantics is used: words have similar meaning if they appear in similar contexts. This translates to, given a corpus of documents, representing a word as a distribution, over the vocabulary, of the frequency of the words that appear in its context in the corpus. These vectors have the limit of being highly dimensional and extrimely sparse: every word would be represented by a vector of tens of thousands dimensions (the size of the vocabulary) the vast majority of which would have value zero.

To solve these problems several techniques, either based on neural networks or matrix factorization, have been proposed [1]. They both start from the sparse and high dimensional co-occurence matrix and obtain a fixed length, dense, and real valued vector for each word. The vectors are not interpretable when taken singularly, but when analyzed and compared to one another they show interesting properties: words that have similar meaning (in the distributional semantics sense) have vectors that are closed to each other according to the cosine or Euclidean distance, and pairs of vectors representing pairs of words analogy, such as man and woman, king and queen, also have the similar distances. Examples of embeddings algorithms with comparable performance based respectively on neural networks and matrix factorization techniques are Word2Vec [2] and GloVe [3].

Event detection is one of the numerous tasks that make use of word embeddings. Part of the complexity of the task stems from the intrinsic ambiguity of what can be defined as an event, the Oxford English Dictionary definition of event "anything that happens, or is contemplated as happening; an incident, occurrence", can be taken as a good baseline. Event detection goes beyond just finding events in unstructured and unprocessed text, it is often concerned with classifying them according to their type and identifying participants and attributes [4].

More recently developed event detection systems make heavy use of deep learning techniques, in particular recurrent neural network, in all of its variants, appears to be the more promising model thanks to the possibility of analyzing arbitrarily long sequences of labled data [5]. Particulary bi-directional long short-term memory network appear to be the preferred model.

Most of the efforts in developing event detection systems are focused on contemporary text and on the biomedical field.

Sample Heading (Third Level) Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

Sample Heading (Fourth Level) The contribution should contain no more than four levels of headings. Table 1 gives a summary of all heading levels.

Table 1. Table captions should be placed above the tables.

Heading level	Example	Font size and style
Title (centered)	**Lecture Notes**	14 point, bold
1st-level heading	**1 Introduction**	12 point, bold
2nd-level heading	**2.1 Printing Area**	10 point, bold
3rd-level heading	**Run-in Heading in Bold.** Text follows	10 point, bold
4th-level heading	*Lowest Level Heading.* Text follows	10 point, italic

Displayed equations are centered and set on a separate line.

$$x + y = z \tag{1}$$

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Fig. 1).

Theorem 1. *This is a sample theorem. The run-in heading is set in bold, while the following text appears in italics. Definitions, lemmas, propositions, and corollaries are styled the same way.*

Proof. Proofs, examples, and remarks have the initial word in italics, while the following text appears in normal font.

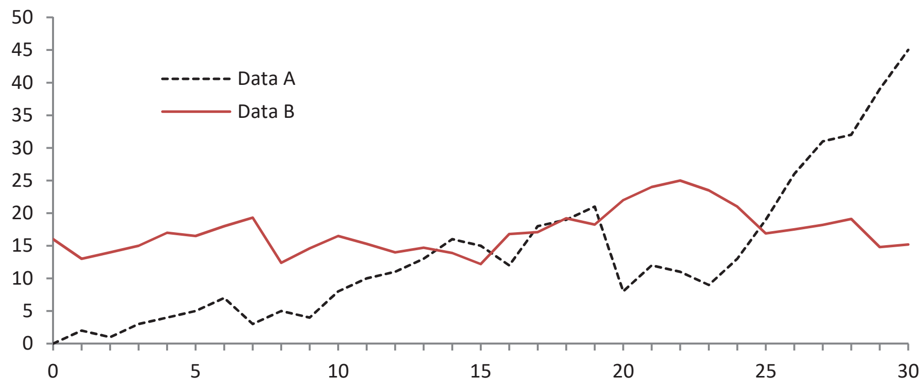

Fig. 1. A figure caption is always placed below the illustration. Please note that short captions are centered, while long ones are justified by the macro package automatically.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal articles [**?**], an LNCS chapter [**?**], a book [**?**], proceedings without editors [**?**], and a homepage [**?**]. Multiple citations are grouped [**?**,**?**,**?**], [**?**,**?**,**?**,**?**].

References

1. F. Almeida and G. Xexo, "Word embeddings: A survey," 2019.
2. T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems 26* (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, eds.), pp. 3111–3119, Curran Associates, Inc., 2013.
3. J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.
4. J. Kalita, "Detecting and extracting events from text documents," 2016.
5. X. Wei and B. Wang, "A survey of event extraction from text," *IEEE Access*, vol. PP, pp. 1–1, 11 2019.