Understanding and Evaluating Medical Concept Embeddings

Andrew L. Beam*, Inbar Freid, Nathan P. Palmer, Isaac S. Kohane
Department of Biomedical Informatics, Harvard Medical School, Boston, MA, 02115, USA
**E-mail: Andrew_Beam@hms.harvard.edu*
www.university_name.edu

Benjamin Kompa
University of North Carolina, Chapel Hill, Chapel Hill, North Carolina ZIP/Zone, USA
E-mail: an_author@laboratory.com

Word embeddings, also known as distributed representations, have seen rapid adoption in natural langauge processing (NLP). Though they are now standard practice in many areas of NLP and machine learning, they are just now beging to attract interest in biomedical informatics. In this article, we present an overview of the existing word embedding methodology and its applicability to biomedical informatics, as well as proposing a set of benchmark for medical concept embedding evaluation. We provide these benchmarks as an R package to the community to encourage quick and easy comparison for new embeddings in the future.

Keywords: Machine Learning; Distributed Representations; Wrord Vectors; Concept Embeddings

1. Introduction

Here is where we will motivate the paper and introduce the key ideas

2. Overview of Word Embeddings

Here is where I will put the overview of exisitng methodology.

3. Benchmarks

Here is where we will put the description of all of the benchmarks, put in `\subsection{}` tags

4. Results

You can obtain these files from the following website: `http://www.wspc.com.sg/style/proceedings_style.shtml`.

5. References

Leave this here for now, I will compile a bibtex file References are to be listed in the order cited in the text in Arabic numerals. BIBTEX users, please use our bibliography style file `ws-procs11x85.bst` for references. Non BIBTEX users can list down their references in the following pattern.

```
\begin{thebibliography}{9}
```

```
\bibitem{jarl88} C. Jarlskog, in {\it CP Violation} (World Scientific,
   Singapore, 1988).

\bibitem{lamp94} L. Lamport, {\it \LaTeX, A Document Preparation System},
   2nd edition (Addison-Wesley, Reading, Massachusetts, 1994).

\bibitem{ams04} \AmS-\LaTeX{} Version 2 User's Guide (American Mathematical
   Society, Providence, 2004).

\bibitem{best03} B.~W. Bestbury, {\em J. Phys. A} {\bf 36}, 1947 (2003).

\end{thebibliography}
```

6. BIBTEXing

If you use the BIBTEX program to maintain your bibliography, you do not use the `thebibliography` environment. Instead, you should include

```
\bibliographystyle{ws-procs11x85}
\bibliography{ws-pro-sample}
```

where `ws-procs11x85` refers to a file `ws-procs11x85.bst`, which defines how your references will look. The argument to `\bibliography` refers to the file `ws-pro-sample.bib`, which should contain your database in BIBTEX format. Only the entries referred to via `\cite` will be listed in the bibliography.

Sample output using `ws-procs11x85` bibliography style file:

BIBTEX database entry type	Sample citation
article	... text.[?,?,?]
proceedings	... text.[?]
inproceedings	... text.[?]
book	... text.[?,?]
edition	... text.[?]
editor	... text.[?]
series	... text.[?]
tech report	See Refs. **?** and **?** for more details
unpublished	... text.[?]
phd thesis	... text.[?]
masters thesis	... text.[?]
incollection	... text.[?]
misc	... text.[?]