

Using the Outlier Detection Task to Evaluate Distributional Semantic Models

Título Using the Outlier Detection Task to Evaluate Distributional Semantic Models

Autores [Pablo Gamallo](#)

Tipo Artículo de revista

Fonte  [Machine Learning and Knowledge Extraction](#), MDPI, Vol. 1, No. 1, pp. 211-223 , 2018.

ISSN 2504-4990

DOI [10.3390/make1010013](#)

Abstract In this article, we define the outlier detection task and use it to compare neural-based word embeddings with transparent count-based distributional representations. Using the English Wikipedia as a text source to train the models, we observed that embeddings outperform count-based representations when their contexts are made up of bag-of-words. However, there are no sharp differences between the two models if the word contexts are defined as syntactic dependencies. In general, syntax-based models tend to perform better than those based on bag-of-words for this specific task. Similar experiments were carried out for Portuguese with similar results. The test datasets we have created for the outlier detection task in English and Portuguese are freely available.

Palabras clave distributional semantics, dependency analysis, evaluation, word similarity

LIGAZÓNS

 [Versión da editorial](#)

DESCARGAS

 [Referencia BibTex](#)

 [Descargar versión do editor](#)

PROXECTOS DE INVESTIGACIÓN

TELEPARES: Tecnologías de la lengua para análisis de opiniones en redes sociales

PROGRAMAS CIENTÍFICOS

Tecnoloxías da Linguaxe Natural