



Luis Nieto Piña

Meaning representation is a central problem in Language technology. By assigning semantic representations to language units such as words or sentences, computer systems are able to integrate meaning into their processes. This is crucial in many of today's language applications such as sentiment analysis or text summarization. Current Machine learning models tend to focus on representing word forms. This might be problematic for words with more than one meaning, given that several meanings are assigned a single representation. Furthermore, these models usually learn semantics from large collections of text, which entails that the word meanings captured depend heavily on the chosen text.

In his doctoral thesis, Luis Nieto Piña presents three models that address those shortcomings. On one hand, the focus is shifted from words to word senses, making it possible to obtain a representation for each meaning of a word. On the other, semantic data is not only obtained from text, but also from lexica to supply the model with curated information about the meanings of words and the relations between them. Luis shows that a combination of these two sources of data yields higher quality word sense representations. In the evaluation of these models, he also demonstrates the utility of these representations, both in established applications and in the development of linguistic resources.

Learning word sense representations from corpora and lexica

Luis Nieto Piña / Splitting rocks:

Splitting rocks: Learning word sense representations from corpora and lexica

Luis Nieto Piña

