编码方式: word2vec 2013 ,放入CNN

模型 CNN LSTM Attention 2017，2018

Bert 2021

Word2vec是Word Embedding 的方法之一，是2013 年由谷歌的 Mikolov提出了一套新的词嵌入方法。在word embedding之前出现的文本表示方法有one-hot编码和整数编码，one-hot编码和整数编码的缺点均在于无法表达词语之间的相似性关系。如何解决这个问题呢？自动学习向量之间的相似性表示，用更低维度的向量来表示每一个单词。