Skip-gram vs CBOW (Continuous Bag-of-Words)

Algorithmically, these models are similar, except that CBOW predicts center words from context words, while the skip-gram does the inverse and predicts source context-words from the center words. For example, if we have the sentence: ""The quick brown fox jumps"", then CBOW tries to predict ""brown"" from ""the"", ""quick"", ""fox"", and ""jumps"", while skip-gram tries to predict ""the"", ""quick"", ""fox"", and ""jumps"" from ""brown"".

Statistically it has the effect that CBOW smoothes over a lot of the distributional information (by treating an entire context as one observation). For the most part, this turns out to be a useful thing for smaller datasets. However, skip-gram treats each context-target pair as a new observation, and this tends to do better when we have larger datasets.

code