Word2vec

Two architectures:

- CBOW (Continuous Bag-of-words):

$$p(w_i | w_{i-h}, \ldots w_{i+h})$$

- Continuous Skip-gram:

$$p(w_{i-h}, \ldots w_{i+h} | w_i)$$

Two ways to avoid softmax:

- Negative sampling
- Hierarchical softmax

Open-source and fast: code.google.com/archive/p/word2vec/