Mooney et al. [1] introduce a method for constructing multiple vector representations of words. Both the methods perform word sense discrimination as a preprocessing step, which can be used to cluster contexts for each word type, making training more expensive. Improvements are suggested in the methods proposed by Neelakantan et al. [3], in which multiple word senses and global representations are computed simultaneously. This is one of the first papers which explore Non Parametric Word Embeddings.

Approach

1. Input: Single sense word embeddings OR Construction of word embeddings [3]
2. Identify top M words for which we compute multiple senses (Generally by frequency)
3. Construct context vectors: Estimate the contextual representation for each word type, completely ignoring sharing such information could yield better results.
4. Perform clustering using the estimate.
5. Use clusters as centers for sense vectors [1]

Further Improvements:

A. Use the estimated sense vectors to assign senses to all occurrences of the words; Retrain using skip gram model.
B. Directly use cluster senses as vectors. The global word vector becomes the average of the sense vectors. (The final vector space remains the same)