Glove paper:

Uses statistics of frequency from cooccurrence matrix.
Uses ratios to better represent meaning linear relations.
Has methods to address overweighting of very frequent words (like the, is) and very rare words by limiting (x/xmax)^alpha and taking weighted least square regression model.
Results depended on vector length, context window size, corpus size and quality etc.
Single epoch model overcome by negative sampling.

Improving Distributional Similarity with Lessons Learned from Word Embeddings:

Positive pointwise mutual information (PPMI) has matrix where rows represent words and columns represent context. Hence, each point represents word-context.
Low dimensional vectors can be made by truncated SVD (singular value decomposition).
In SGNS (skipgram with negative sampling), dimensional vector represents both word and context.
3 types of hyper parameters:

1) Preprocessing - dynamic context window, handling rare words.
2) Association metric - negative sampling, smoothing context distribution.
3) Post processing - adding context vectors, symmetric vs non symmetric variants of SVD, vector normalisation (rows, columns, none, both).

No model was consistently superior to others. So, hyper parameter tuning needs to be considered before we decide on best model. Also, in some cases, hyper parameter tuning is more beneficial than expanding dataset to get better results, helping in computation.
Context distribution smoothing works consistently and do not use SVD correctly.

Evaluation methods for unsupervised word embeddings:

At present, extrinsic and intrinsic evaluations of word embeddings- final task vs specific tests.
Diverse dataset is required based on frequency of words, parts of speech and abstractness/ concreteness of words.
Metric aggregation is difficult since scores and distances vary greatly in embedded space. Even ranking scores is not good since how do we know if (dog, cat) is better than (apple, banana).
Extrinsic evaluations favour various embeddings for various tasks. At the same time, they are not properly reflective of intrinsic evaluations.
Frequent words seems to have lot of influence on cosine similarity and word's nearest neighbour. Frequent words are coming as close neighbours for lot of words. As such, further efforts are needed to separate influence of frequent words both in evaluation of embeddings as well as in creation of vector space since vector space needs to reflect semantic relationships and not frequency.

A Latent Variable Model Approach to PMI-based Word Embeddings:

Not understood.

Linear Algebraic Structure of Word Senses, with Applications to Polysemy:

Polysemous words are being described through linear superposition in modern papers instead of multiple vectors for multiple meanings. Sparse coding helps in this regard. This is shown in paper theoretically. Multiple meanings are linearly superpositioned.

On the Dimensionality of Word Embedding:

Optimal vector dimensionality can be selected based on bias-variance tradeoff of pairwise inner product (PIP) loss (theoretical framework). Pip loss is minimised.