Name	Description	Representations	Advantages and Disadvantages	Relationships to Other Models	References
Vector space model	- records the occurrence of words in documents in a matrix and uses the cosine to measure similarity - a number of weighting schemes have been tried to emphasize words that occur in few contexts, log entropy scheme has been found to be effective	- words are represented by a set of weights across documents - documents are represented as a set of weights across words - has no representation of topics or concepts	- simple to implement and use - scales well - surprizingly effective - does not take advantage of latent structure in a corpus - not easily interpretable	- this is the foundational model on which most of the other models are derived	G. Salton , A. Wong , C. S. Yang, A vector space model for automatic indexing, Communications of the ACM, v.18 n.11, p.613-620, Nov. 1975.
Latent Semantic Analysis	- takes a weighted occurrence matrix - applies singular value decomposition - reduces the dimensionality of the matrix – typically down to about 300 dimensions	- words are represented as sets of factor loadings - documents are represented as sets of factor loadings - has no representation of topics or concepts	- once space has been constructed it is very efficient - words with similar meanings come to be represented by similar vectors, it does take advantage of latent structure - while it has no immediate topic or concept representation, SVD solution can be rotated to make interpretable dimensions - has been shown to be practically useful in a wide variety of domains	- starts with the same word by document matrix as vector space model	Kintsch, W., McNamara, D., Dennis, S. & Landauer, T. (2006). Handbook of Latent Semantic Analysis. Mahwah:NJ. Lawrence Erlbaum Associates. (especially Chapter 2).
Topics model	- assumes a simple Bayesian generative model of a document - infers the posterior probability of words given each of a set of topics using a Gibbs sampler	- words are probability distributions over topics - documents are probability distributions over topics - topics tend to be interpretable	- unlike LSA, the topics model respects the fact that the corpus provides a set of counts for words in documents - the Gibbs sampler is quite efficient and easy to implement - topics tend to be interpretable - by running the Gibbs sampler over a given	- the Latent Dirichlet Allocation method of Blei et al uses the same generative model but employs EM as the inference method rather than a Gibbs sampler	Griffiths, T. L., & Steyvers, M. (2002). A probabilistic approach to semantic representation. In C. D. Schunn & W. D. Gray (Eds.), Proceedings of the 24th Annual Conference of the Cognitive Science Society: Lawrence Erlbaum Associates.

Name	Description	Representations	Advantages and Disadvantages	Relationships to Other Models	References
			document it is possible to build a meaning representation for a word within a document as opposed to a general meaning representation, so polysemous words receive different representations depending on context - the Bayesian nature means that words start out being different and can only become similar given sufficient evidence, whereas with LSA words can be very similar given a single occurrence		Latent Dirichlet Allocation, David Blei, Andrew Y. Ng and Michael Jordan. Journal of Machine Learning Research, 3:993-1022, 2003.
Independent Components Analysis (ICA)	- starts with document representations of words as per vector space models - defines a set of similar words, neutral words and dissimilar words - representations updates by adding similar-neutral and subtracting dissimilar minus neutral	- can produce both word and document representations (using same procedure)	- in our experience setting the boundaries for what is similar and dissimilar is critical and non trival - this method does not seem to have been widely adopted	- derivative of vector space model	Isbell, C. L. and Viola, P.: Restucturing sparse high dimensional data for effective retrieval, In: Advances in Neural Injbrmation Processing Systems 11, 1998, pp. 480486.
Sparse ICA	- take document representations of words and cluster (using K means, fuzzy C means etc) - take centroid of each cluster - word representation is the loadings on these centroid vectors - takes advantage of the	- produces word and document vectors - centroid vectors can be thought of as topics	- quite efficient and produces topic like vectors - does not have the ability to predicate meanings like the topics model	- derived from vector space model	A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, Y. Y. Zeevi, Sparse ICA for blind separation of transmitted and reflected images, Intl. Journal of Imaging Science and Technology (IJIST), 15(1), pp. 84-91, 2005.

Name	Description	Representations	Advantages and Disadvantages	Relationships to Other Models	References
	fact that in sparse systems independent components are largely exclusive				
Nonnegative Matrix Factorization	- decompose raw occurrence matrix into positive matrices U and V using a gradient descent method - a variable learning rate is used which allows updates to be multiplicative and ensures nonnegativity	- produces word and document reprentations - also produces something similar to topic representations	- easy to implement and efficient - produces interpreatable dimensions	- has been shown to be equivalent to Latent Dirichlet Allocation under certain conditions	Xu, W., Liu, X., & Gong, Y. (2003). Document clustering based on non-negative matrix factorization. SIGIR, Toronto, Canada.
Syntagmatic Paradigmatic Model	- estimate word probability distribution for each word in a sentence (the paradigmatic associates) given the surrounding words (the syntagmatic associates) - bind each word to the probability distribution of the associated slot using the outer product and add across the sentence to form a relational representation of the sentence	- produces word and sentence representations	- unlike other models produces relational representations of sentences and uses word order information - computationally more expensive than other methods	- while other models are all to some extent derivatives of the vector space model the SP model has quite a different purpose and mechnaism	Dennis, S. (2004). An unsupervised method for the extraction of propositional information from text. Proceedings of the National Academy of Sciences, 101, 5206-5213. Dennis, S. (2005). A Memory-based Theory of Verbal Cognition. Cognitive Science. 29(2). 145-193.

Note: References provide mathematical treatments of each of the models.