## 4.4 Document Retrieval

We used 20-newsgroup and Reuters datasets to evaluate model performance on a document retrieval task. To decide whether a retrieved document is relevant to the query document, we simply check if they have the same class label. This is the only time that the class labels are used. For the Replicated Softmax, the mapping from a word-count vector to the values of the latent topic features is fast, requiring only a single matrix multiplication followed by a componentwise sigmoid non-linearity. For the LDA, we used 1000 Gibbs sweeps per test document in order to get an approximate posterior over the topics. Figure 3 shows that when we use the cosine of the angle between two topic vectors to measure their similarity, the Replicated Softmax significantly outperforms LDA, particularly when retrieving the top few documents.

## 5 Conclusions and Extensions

We have presented a simple two-layer undirected topic model that be used to model and automatically extract distributed semantic representations from large collections of text corpora. The model can be viewed as a family of different-sized RBM’s that share parameters. The proposed model have several key advantages: the learning is easy and stable, it can model documents of different lengths, and computing the posterior distribution over the latent topic values is easy. Furthermore, using stochastic gradient descent, scaling up learning to billions of documents would not be particularly difficult. This is in contrast to directed topic models, where most of the existing inference algorithms are designed to be run in a batch mode. Therefore one would have to make further approximations, for example by using particle filtering [3]. We have also demonstrated that the proposed model is able to generalize much better than LDA in terms of both the log-probability on held-out documents and the retrieval accuracy.

In this paper we have only considered the simplest possible topic model, but the proposed model can be extended in several ways. For example, similar to supervised LDA [1], the proposed Replicated Softmax can be easily extended to modeling the joint the distribution over words and a document label, as shown in Fig. 4, left panel. Recently, [11] introduced a Dirichlet-multinomial regression model, where a prior on the document-specific topic distributions was modeled as a function of observed metadata of the document. Similarly, we can define a conditional Replicated Softmax model, where the observed document-specific metadata, such as author, references, etc., can be used...