"Pop"	*"Indie"*	*"Jazz"*	*"Classical"*	*"Metal"*	*"Reggae"*	*"Electronic"*	*"Experimental"*	*"Country"*
pop	indie	chillout	piano	metal	reggae	house	instrumental	country
female vocal	rock	lounge	instrumental	death metal	funk	electro	ambient	classic country
dance	alternative	chill	ambient	thrash metal	funky	electronic	experimental	male vocal
electronic	indie rock	downtempo	classic	brutal death metal	dance	dance	electronic	blues
sexy	post punk	smooth jazz	beautiful	grindcore	hip-hop	electric house	psychedelic	folk
love	psychedelic	relax	chillout	heavy metal	party	techno	progressive	love songs
synth pop	new wave	ambient	relax	black metal	sexy	minimal	rock	americana

Table 3. Top 7 tags from 9 latent factors for PMF-Stoc-full with $J = 512$. For each factor, we assign the closest music genre on top. As is evident, each factor corresponds to a particular aspect of a music genre.

can be used to infer the most appropriate tags given the audio alone. The Poisson model is naturally less sensitive to zero values than some alternatives, making it a good match to "noisy" training examples derived from real users' taggings, where the fact that no user has applied a tag does not necessarily imply that the term is irrelevant. By learning this model using stochastic variational inference, we are able to efficiently exploit much larger training sets than are tractable using batch approaches, making it feasible to learn from an entire set of over 370k tagged examples. Although much of the improvement comes in the earlier iterations, we see continued improvement implying this approach can benefit from much larger, effectively unlimited sources of tagged examples, as might be available on a commercial music service with millions of users.

There are a few areas where our model can be easily developed. For example, stochastic variational inference requires we set the learning rate parameters t_0 and κ, which is application-dependent. By using adaptive learning rates for stochastic variational inference [13], model inference can converge faster and to a better local optimal solution. From a modeling perspective, currently the hyperparameters for weights $\boldsymbol{\theta}$ are fixed, indicating that the sparsity level of the weight for each song is assumed to be the same *a priori*. Alternatively we could put *song-dependent* hyper-priors on the hyperparameters of $\boldsymbol{\theta}$ to encode the intuition that some of the songs might have denser weights because more tagging information is available. This would offer more flexibility to the current model.

7. ACKNOWLEDGEMENTS

The authors would like to thank Matthew Hoffman for helpful discussion. This work was supported in part by NSF grant IIS-1117015.

8. REFERENCES

[1] Last.fm dataset, the official song tags and song similarity collection for the Million Song Dataset. `http://labrosa.ee.columbia.edu/millionsong/lastfm`.

[2] T. Bertin-Mahieux, D. Ellis, B. Whitman, and P. Lamere. The Million Song Dataset. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 591–596, 2011.

[3] L. Bottou. Online learning and stochastic approximations. *On-line learning in neural networks*, 17(9), 1998.

[4] A. T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009.

[5] D. Eck, P. Lamere, T. Bertin-Mahieux, and S. Green. Automatic generation of social tags for music recommendation. In *Advances in Neural Information Processing Systems*, pages 385–392, 2007.

[6] K. Ellis, E. Coviello, A. Chan, and G. Lanckriet. A bag of systems representation for music auto-tagging. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21(12):2554–2569, 2013.

[7] M. Hoffman, D. Blei, and P. Cook. Easy as CBA: A simple probabilistic model for tagging music. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 369–374, 2009.

[8] M. Hoffman, D. Blei, and P. Cook. Bayesian nonparametric matrix factorization for recorded music. In *Proceedings of the 27th Annual International Conference on Machine Learning*, pages 439–446, 2010.

[9] M. Hoffman, D. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

[10] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

[11] D. Liang, M. Hoffman, and D. Ellis. Beta process sparse nonnegative matrix factorization for music. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 375–380, 2013.

[12] J. Paisley, D. Blei, and M.I. Jordan. Bayesian nonnegative matrix factorization with stochastic variational inference. In E.M. Airoldi, D. Blei, E.A. Erosheva, and S.E. Fienberg, editors, *Handbook of Mixed Membership Models and Their Applications*. Chapman and Hall/CRC, 2015.

[13] R. Ranganath, C. Wang, D. Blei, and E. Xing. An adaptive learning rate for stochastic variational inference. In *Proceedings of The 30th International Conference on Machine Learning*, pages 298–306, 2013.

[14] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*, pages 1257–1264, 2008.

[15] D. Tingle, Y.E. Kim, and D. Turnbull. Exploring automatic music annotation with acoustically-objective tags. In *Proceedings of the international conference on Multimedia information retrieval*, pages 55–62. ACM, 2010.

[16] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech and Language Processing*, 16(2):467–476, 2008.

[17] B. Xie, W. Bian, D. Tao, and P. Chordia. Music tagging with regularized logistic regression. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 711–716, 2011.