In order to avoid the loss of information during the compression phase of the embedding based approaches, tree-based methods have been proposed that try to partition the label space similar to a decision tree. It recursively partitions the huge label space in subtrees until only a few labels are left at each leaf node. A base classifier at each leaf node then focuses on only the active labels in the node. The LPSR [31] method focuses on learning a hierarchy over a base classifier or ranker starting with a base multi-label classifier for the entire label set - this becomes computationally expensive to train if a discriminative classifier (e.g. SVM) is used. Instead of using a base classifier, MLRF [1] uses an ensemble of randomized trees with a modified Gini index for partitioning the nodes. In FastXML [21] an NDCG-based objective is used at each node of the hierarchy for optimization.

Despite the success of deep learning in many fields, it has not been explored much for XMLC tasks. Recently, a CNN based approach (XML-CNN [18]) has been proposed, which uses convolutional layers for text representation and a feed forward layer acting as a bottleneck layer for scalability. This has been shown to outperform both embedding based and tree based approaches for XMLC, therefore we chose this method for comparison in our experiments.

6 CONCLUSION

We have addressed cold start thread recommendation in online forums, which is an important task to ensure user engagement. We recommend newly-posted threads to interested users in the community for participation. Mainstream recommendation systems cannot use collaborative filtering to address this phenomenon as there is no interaction history for such items.

We have applied an alternative approach utilizing extreme multi-label classification. In particular, we proposed a novel neural network architecture consisting of stacked bi-directional GRUs for text encoding, coupled with cluster-sensitive attention to address scalability, and sparsity.

Specifically, leveraging our insight that sets of users display different levels of interest within a long post text, the cluster-sensitive attention incorporates user interests by learning multiple attention layers for attending to different parts of a text. This cluster-sensitive attention layer also helps us in addressing the sparsity issues usually associated with extreme multi-label classification approaches, by exploiting the correlation between users within clusters. Thorough experimental evaluation show that the proposed model outperforms existing content based recommendation systems, deep learning based text classification systems, as well as state-of-the-art multi-label classification approaches.

In the future, we plan to model how interests of community members change over time. Not all users will remain interested in the same topic over a long course of time as their experiences and expertise change. Also, we encourage the research community to try our approach in other domains where recommending new items to interested users is the priority such as news articles, tweet recommendation, social media news feed generation and so on.

REFERENCES

[1] Rahul Agrawal, Archit Gupta, Yashoteja Prabhu, and Manik Varma. 2013. Multi-label learning with millions of labels: Recommending advertiser bid phrases for web pages. In *Proc. of WWW*. ACM, 13–24.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[3] Krishnakumar Balasubramanian and Guy Lebanon. 2012. The landmark selection method for multiple output prediction. In *Proc. of ICML*. Omnipress, 283–290.

[4] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks* 5, 2 (1994), 157–166.

[5] Kush Bhatia, Himanshu Jain, Purushottam Kar, Manik Varma, and Prateek Jain. 2015. Sparse local embeddings for extreme multi-label classification. In *Proc. of NIPS*. 730–738.

[6] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[7] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *Proc. of NIPS Deep Learning and Representation Learning Workshop* (2014).

[8] Moustapha M Cisse, Nicolas Usunier, Thierry Artieres, and Patrick Gallinari. 2013. Robust bloom filters for large multilabel classification tasks. In *Proc. of NIPS*. 1851–1859.

[9] Kishaloy Halder, Min-Yen Kan, and Kazunari Sugiyama. 2017. Health Forum Thread Recommendation Using an Interest Aware Topic Model. In *Proc. of CIKM*. ACM, 1589–1598.

[10] Daniel J Hsu, Sham M Kakade, John Langford, and Tong Zhang. 2009. Multi-label prediction via compressed sensing. In *Proc. of NIPS*. 772–780.

[11] Ashish Kapoor, Raajay Viswanathan, and Prateek Jain. 2012. Multilabel classification using bayesian compressed sensing. In *Proc. of NIPS*. 2645–2653.

[12] Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proc. of EMNLP*.

[13] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014).

[14] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proc. of SIGKDD*. ACM, 426–434.

[15] Daniel D Lee and H Sebastian Seung. 2001. Algorithms for non-negative matrix factorization. In *Proc. of NIPS*. 556–562.

[16] Xiaopeng Li and James She. 2017. Collaborative variational autoencoder for recommender systems. In *Proc. of SIGKDD*. ACM, 305–314.

[17] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. 2017. A structured self-attentive sentence embedding. *ICLR* (2017).

[18] Jingzhou Liu, Wei-Cheng Chang, Yuexin Wu, and Yiming Yang. 2017. Deep Learning for Extreme Multi-label Text Classification. In *Proc. of SIGIR*. ACM, 115–124.

[19] Andriy Mnih and Ruslan R Salakhutdinov. 2008. Probabilistic matrix factorization. In *Proc. of NIPS*. 1257–1264.

[20] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation.. In *Proc. of EMNLP*.

[21] Yashoteja Prabhu and Manik Varma. 2014. Fastxml: A fast, accurate and stable tree-classifier for extreme multi-label learning. In *Proc. of SIGKDD*. ACM, 263–272.

[22] Ruslan Salakhutdinov and Andriy Mnih. 2008. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proc. of ICML*. 880–887.

[23] Erez Shmueli, Amit Kagian, Yehuda Koren, and Ronny Lempel. 2012. Care to Comment?: Recommendations for Commenting on News Stories. In *Proc. of WWW*. ACM, 429–438.

[24] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In *Proc. of NIPS*. 2643–2651.

[25] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research* 11, Dec (2010), 3371–3408.

[26] Maksims Volkovs, Guangwei Yu, and Tomi Poutanen. 2017. DropoutNet: Addressing Cold Start in Recommender Systems. In *Proc. of NIPS*. 4964–4973.

[27] Chong Wang and David M Blei. 2011. Collaborative topic modeling for recommending scientific articles. In *Proc of SIGKDD*. ACM, 448–456.

[28] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative Deep Learning for Recommender Systems. In *Proc. of SIGKDD*. ACM, 1235–1244.

[29] Hao Wang, SHI Xingjian, and Dit-Yan Yeung. 2016. Collaborative recurrent autoencoder: recommend while learning to fill in the blanks. In *Proc. of NIPS*. 415–423.

[30] Xinxi Wang and Ye Wang. 2014. Improving Content-based and Hybrid Music Recommendation Using Deep Learning. In *Proceedings of the 22Nd ACM International Conference on Multimedia*. 627–636.

[31] Jason Weston, Ameesh Makadia, and Hector Yee. 2013. Label partitioning for sublinear ranking. In *Proc. of ICML*. 181–189.

[32] Bin Xu, Jiajun Bu, Chun Chen, and Deng Cai. 2012. An Exploration of Improving Collaborative Recommender Systems via User-item Subgroups. In *Proc. of WWW*. ACM, 21–30.

[33] Yi Zhang and Jeff Schneider. 2011. Multi-label output codes using canonical correlation analysis. In *Proc. of AISTATS*. 873–882.