

Learning Reliability Evaluation Models of Power Communication Network Equipments with Capsule Networks

Haiyang Wu, Xiao Lu, Dawei Su
State Grid Jiangsu Electric Power Co., Ltd.
Nanjing, China
wuhaiyang@foxmail.com
lx_sz@js.sgcc.com.cn
david_su@js.sgcc.com.cn

Weiwei Miao, Li Zhang, Qi Sun
Inf. and Comm. Branch
State Grid Jiangsu Electric Power Co., Ltd.
Nanjing, China
wwmiao@js.sgcc.com.cn
zhang@js.sgcc.com.cn
amysun@js.sgcc.com.cn

Abstract—To evaluate reliability of equipments plays a key role for the robustness of modern power communication network. Such tasks imply to learn a regression model with limited samples and heterogeneous formats. In this paper, we address this problem by introducing a deep learning based model containing the latest proposed capsule structures. The model can take raw maintenance data as input directly by treating them as unified document without any extra manual preprocessing. We propose a multiple stages strategy to train the model with original data as well as generated perturbed data from a decoder as augmentation. Experimental results demonstrate that not only the proposed method shows acceptable performance to predict the reliability of communication equipments, but also it shows potentiality especially in learning deep models with fewer samples in different Natural Language Processing tasks. Thus, the proposed capsule networks framework with data generation mechanism could be considered as a promising way to drive deep models in practical learning tasks in which only limited training data is available.

Keywords—Capsule Networks, Natural Language Processing, Deep Learning, Smart Grid, Reliability Evaluation

I. INTRODUCTION

Recently, electric power communication networks have experienced a significant development from the perspective of equipments and coverage area. The modern communication network is important infrastructure of power system since not only it provides the service of network management, but also it delivers pivotal running data of various equipments and sensors on power grid. Therefore, the reliability evaluation and analysis of electric power communication network itself also becomes increasingly important. Conventional power communication network maintenance normally corresponds to periodically inspections or fault recovery after alarm [1]. To evaluate the reliability status and realize real-time prediction for equipment faults based on running records gradually extract attention from the community since it provides a potential way to protect networks from serious faults before alarm occurs [2].

Learning reliability evaluation model for equipments of power communication network is still a challenging problem

since it requires extracting mutual information from multimedia data with complicated structures. Conventional methods for sensor health prediction [3] or equipment reliability evaluation [2] normally rely on heuristically designed features from raw data as well as training classifiers such as Support Vector Machine (SVM) [4], decision tree [5]. The disadvantages of such frameworks are obvious. The human designed features can only consider local information on single node, rather than treating equipments on the same network comprehensively. The training of classifier is not end-to-end, thus the prediction performance of resulting model is strongly restricted by the quality of feature extraction algorithm, which mainly depends on the specific knowledge of developer. On the other side, raw running data of power communication networks normally contains heterogeneous formats such as documents and structural data. It is difficult to mine relevant information manually as preprocessing for further regression. Hence in this paper we propose to learn the evaluation model based on capsule networks [6] by treating running data as documents with unified format directly.

Deep neural networks based methods have achieved extensive applications in the fields of Natural Language Processing (NLP) [7], computer vision [8], etc. The end-to-end training framework ensures that such models can autonomously learn the most representative features for specific task. However, the huge quantity of parameters within such deep models normally require large scale data to drive training. It is not a serious problem for daily NLP tasks due to easily accessible training corpus. Contrarily, for the problem of equipments reliability evaluation, the training set is relatively small and most data shows similar pattern. Differing from typical deep learning based models, capsule network integrates unsupervised routing with standard back propagation (BP) learning of neural networks. It helps model to extrapolate latent samples with the same distribution by affine transformation, and thus significantly prompt the generalization capacity especially on small datasets [9], [10].

In this paper, we proposed a simple generation framework to further reduce the amount of samples for learning a capsule network for NLP tasks. On top of that, we apply the proposed model on the problem of reliability evaluation of power communication network equipments.

Our proposed method outperforms conventional regression methods with much simpler preprocessing mechanisms.

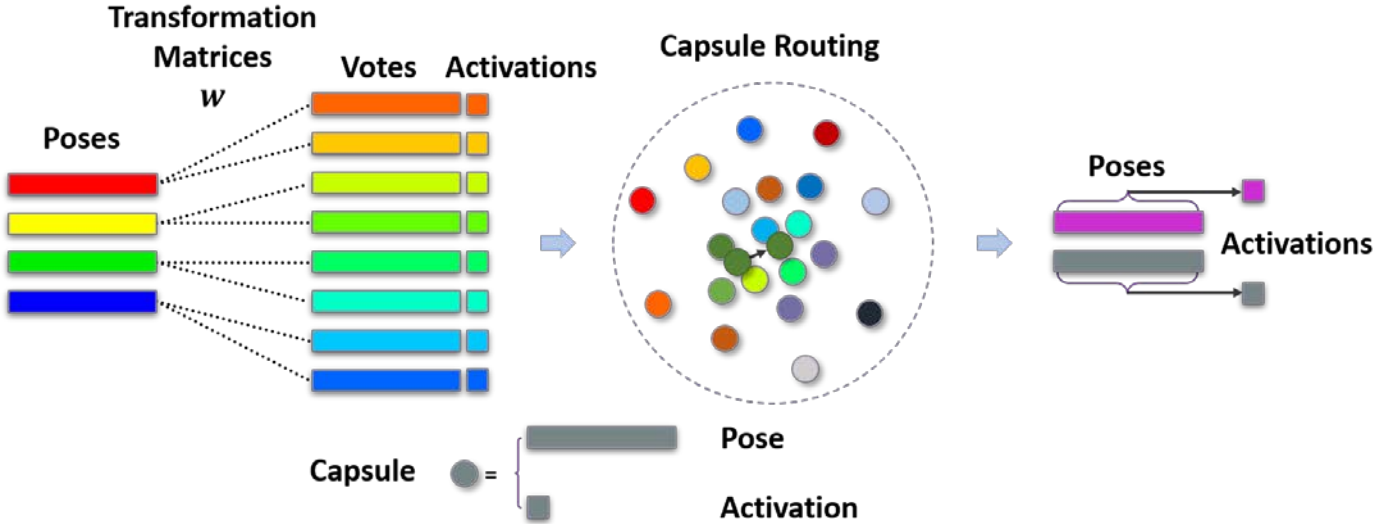


Fig. 1. Capsules and dynamic routing between them. Here 4 capsules are affine transformed to 8 candidates and then routed to 2 capsules as output.

II. CAPSULE NETWORKS

Capsule structures as well as routing algorithm is proposed in [11]. As illustrated in Fig. 1, the building block here is so called capsule consisting of a vector as pose and a corresponding scalar as activation. The pose vector, normally 8D or 16D, is realized by grouping features from previous layers of neural network. A scalar is combined to pose as the nonlinear activation of the capsule. As the output of network, activations represent the magnitude of capsule corresponding to specific semantic category. Meanwhile, the pose of capsule is exploited to reconstruct instantiation information of input as extra regularization for learning.

Based on the capsule structures, a dynamic routing algorithm can be exploited to compress capsules and to formulate the relationship over different capsules. At the first stage of routing, input poses of capsules are extrapolated by affine transformation matrices for more abundant information. Then some unsupervised clustering methods can be deployed on transformed candidates. Here routing methods are only based on the similarity between capsules without any supervision. Capsule networks integrate conventional discriminative BP learning with generative affine transformation and routing, thus demonstrate some attractive features. The output of capsule networks can serve as conventional probability of hypothesis while pose vector can bring specific interpretable information as auxiliary evidence to final decision. Also, since the extrapolation during affine transformation and generative routing method, capsule network can be learned with many fewer samples and prevail deep models [9], [10].

A. Routing by Kernel Density Estimation

The routing method originally proposed in [6] reweighs the connections between candidate capsules and output capsules by a heuristic clustering method with dot product as the metric. The clustering ensures that similar capsules contribute to the same resulting capsule as the representative feature. However, the so called “routing-by-agreement” framework in [6] requires too intensive computation during routing. In this paper, we adopt the kernel density estimation

(KDE) based clustering as a fast routing algorithm [12].

Given a defined distance metric $d(\mathbf{u} - \mathbf{v})$ between pose of candidate capsule \mathbf{u} and pose of output cluster \mathbf{v} , the KDE based routing aims to maximize the weighted sum of density estimations at clusters as

$$\hat{f}(\mathbf{v}, \mathbf{r}) = \frac{1}{n_l z_k} \sum_{j=1}^{n_{l+1}} \sum_{i=1}^{n_l} r_{ij} a_i^u k(d(\mathbf{v}_j - \mathbf{u}_i)), \quad (1)$$

where n_l and n_{l+1} are the number of capsules at input and output layers respectively, z_k is a partition constant and $k(\cdot)$ is the kernel function. r_{ij} s are weights of connections between capsules and clusters. a_i^u is the activation of input capsule.

To minimize $\hat{f}(\mathbf{v}, \mathbf{r})$ in (1), the clusters \mathbf{v}_j s and weights r_{ij} s are optimized alternately. At the τ -th iteration, given fixed r_{ij}^τ , new clusters $\mathbf{v}_j^{\tau+1}$ s can be delivered as from old \mathbf{v}_j^τ s.

$$\mathbf{v}_j^{\tau+1} = \frac{\sum_{i=1}^{n_l} r_{ij}^\tau a_i^u k'(d(\mathbf{v}_j^\tau - \mathbf{u}_i)) \mathbf{u}_i}{\sum_{i=1}^{n_l} r_{ij}^\tau a_i^u k'(d(\mathbf{v}_j^\tau - \mathbf{u}_i))}, \quad (2)$$

On the other hand, r_{ij} is optimized by standard gradient descent method as

$$r_{ij}^{\tau+1} = r_{ij}^\tau + \alpha a_i^u k(d(\mathbf{v}_j^\tau - \mathbf{u}_i)), \quad (3)$$

where α is a constant to control the step of gradient descent. The whole entirety of KDE based routing can be summarized as algorithm 1.

Algorithm 1 Dynamic routing based on mean shift.

Require: poses \mathbf{u}_i , activations a_i^u

 Initialize $\forall i, j: r_{ij} = 1 / n_{i+1}$
for τ iterations **do**

$$1. \forall i, j: r'_{ij} \leftarrow \frac{r_{ij}}{\sum_j r_{ij}}$$

$$2. \forall j: \mathbf{v}_j \leftarrow \frac{\sum_i r'_{ij} a_i^u k'(d(\mathbf{v}_j - \mathbf{u}_i)) \mathbf{u}_i}{\sum_i r'_{ij} a_i^u k'(d(\mathbf{v}_j - \mathbf{u}_i))}$$

$$3. \forall i, j: r_{ij} \leftarrow r_{ij} + a_i^u k(d(\mathbf{v}_j - \mathbf{u}_i))$$

end for
return capsules with poses \mathbf{v}_j

Here α is simply set to 1. Given the resulting poses \mathbf{v}_j , nonlinear activations are attached as

$$a_j^v = \left(\sum_{i=1}^{n_j} r'_{ij} a_i^u \left(k \left(\sum_{d=1}^D d(u_{id} - \beta_{jd} v_{jd}) + \beta_{j0} \right) \right) \right), \quad (4)$$

where r'_{ij} is the normalized version of r_{ij} as in step 1 of algorithm 1 and D is the dimension of pose vector.

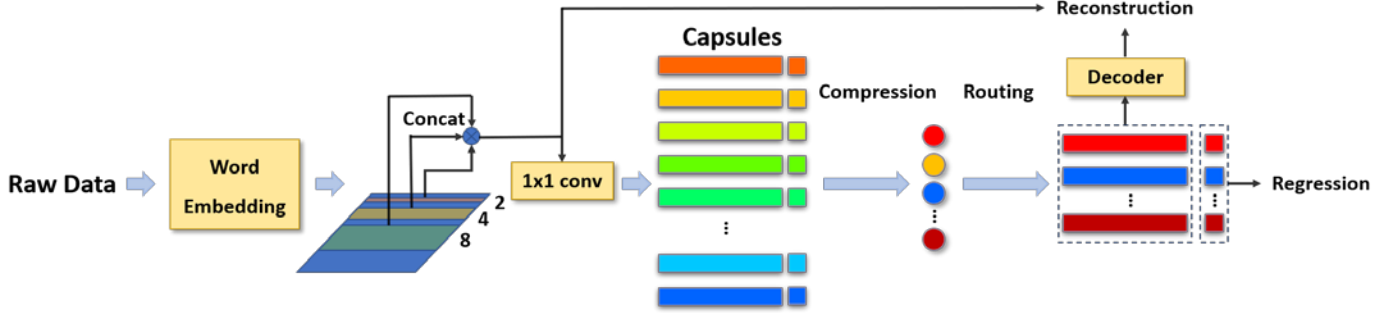


Fig. 2. Evaluation framework based on capsule networks.

Then the dynamic routing algorithm as depicted in Fig. 1 is executed to further combine these compressed capsules as resulting representative capsules. The activations of representative capsules are trained for regression. Simultaneously, poses of capsule are exploited to reconstruct features of input instance by a decoder.

Here we adopt a structure consisting of one fully connected layer and three deconvolutional layers as decoder to reconstruct instantiation features from poses.

The decoder not only helps to regulate the learning as proposed in [6], but also it can generate extra samples sharing similar distribution with original samples. As shown in Fig. 3, at the first stage of training, both capsules network and decoder are trained from original data. Then at the second stage, we generate extra data by feeding original data with perturbation into the decoder learned at the first stage. Here we adopt the proposed perturbation method in [10] at each dimension of the convolutional features of input data. Finally we re-train the capsule network with the combination of both original data and perturbed data at the third stage. Our experimental results demonstrate that the augmentation method based on the decoder can help to drive the learning of model with relatively fewer samples as well as trivial reduction of performance. This is especially suitable in the

$\beta_j \in \mathbb{R}^{D+1}$ is linear coefficient at each entry of pose.

III. RELIABILITY EVALUATION FRAMEWORK

As illustrated in Fig. 2, our proposed evaluation framework treats input data as natural documents with unified structure. The input is processed by word embedding method to get vectors with the same dimension. Then convolutional filters with different sizes (2, 4 and 8) are deployed to extract features from embedded words. Features are concatenated directly to formulate the relationship between data at different scales. A 1×1 convolutional filter is implemented at each position of features to form the poses of capsules. For every pose, an activation as shown in Fig. 2 is attached as the softmax version of the length of pose.

Since the amount of capsules now is too much to realize routing with a practical computation cost, capsules are compressed in the format of weighted sum as

$$\hat{\mathbf{u}}_k = \sum_i b_{ki} \mathbf{u}_i, \quad (5)$$

where b_{kj} is the weight learned by standard BP. By compression the number of capsules is reduced to a reasonable level for routing and some outliers are removed.

scenario of learning tasks with limited training samples, such as reliability evaluation of power communication network equipments.

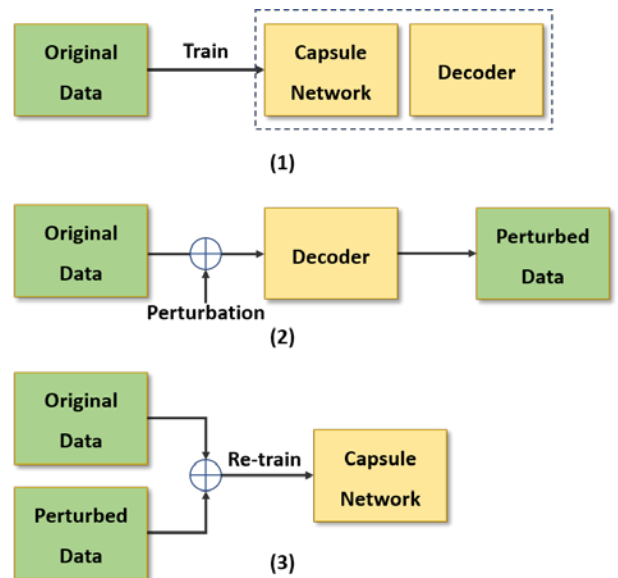


Fig. 3. Three stages for learning capsule network with few samples and augmentation method.

IV. EXPERIMENTAL RESULTS

A. Text Classification with Fewer Samples

Our main concern in this paper is to learn evaluation models with fewer samples. To this end, we first validate the proposed three stages training strategy on a typical multilabel text classification task. The experiments are executed on the dataset EUR-Lex [13], which consists of 3,956 labels and 15.59 examples per label. By varying the fraction of training samples from 30% to 100%, we study the influence brought by extra perturbed data that are generated by decoder. The final performances of models are compared in Fig. 4.

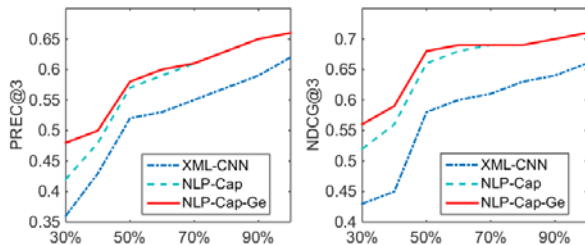


Fig. 4. Performances of models on EUR-Lex with different number of training samples.

Here we compared three methods: XML-CNN [14], NLP-Cap [9] and our proposed NLP-Cap with perturbed samples (NLP-Cap-Ge). The performances at two standard rank-based tasks, Precision@k and Normalized Discounted Cumulative Gain (NDCG@k), are considered. From Fig. 4, one can observe that the proposed method effectively exploited the generation capability of the decoder. The complementary samples during training prompt the performance of model with obvious margin especially in the cases with insufficient training samples. In contrast, when training samples are abundant, since our proposed method is basically similar to NLP-Cap method, both of them show approximately the same performance. The potentiality of sample generation is specifically suitable for learning tasks with fewer samples, such as reliability evaluation of equipments.

B. Reliability Evaluation with Capsule Network

In this section, we empirically test the capacity of capsule networks to evaluate the reliability of power communication network equipments. The data is collected from operation and maintenance data of optical transmission equipments in state grid. The training date covers the range from 2016 to 2017, while we adopt the data at 2018 as test set. We treat data within one week as the essential unit of sample. Each sample is fed into our proposed evaluation framework as Fig. 2 as input with unified document format. The information of data mainly includes both specific features of equipments such as optical power, jitter, drift, Bit Error Rate (BER) and Signal Noise Rate (SNR), etc., and environment parameters such as temperature and power supply state of equipment room. To evaluate the reliability of equipment, we propose to comprehensively consider following features of an equipment: number of interface alarms, number of power alarms, mean of CPU temperature, bias current and receive/launch optical power. All these features within one week are weighed by empirical weights and finally normalized to $[0, 1]$ as resulting evaluation, i.e., reliability of the equipment.

Since our average training data for every equipment is less than 1000 samples. We adopt the training strategy with generated data in Fig. 3 for higher generalization. Performances of four different methods are listed in Table I. First, we train a CNN with similar architecture as Fig. 2, except that capsule layer is replaced by two fully connected layers. One can see that the training cannot converge to a rational point since samples are insufficient for normal deep models. Then we learn a conventional capsule network with only the first stage in Fig. 3. The final regression error is significantly reduced. For comparison, we also learn a linear SVM based on features extracted by conventional layers in the left part of Fig. 2. One can see that the evaluation result is also unacceptable as direct CNN model. Finally, capsule network with generated samples performs best here.

TABLE I. FOUR METHODS ARE COMPARED ON THE TASK OF RELIABILITY REGRESSION. WE REPORT MEAN ERROR BETWEEN INFERENCE RELIABILITIES AND THE GROUND TRUTH.

| Method | Regression error |
|--------|------------------|
| CNN | 0.43 |
| Cap | 0.09 |
| SVN | 0.41 |
| Cap-Ge | 0.03 |

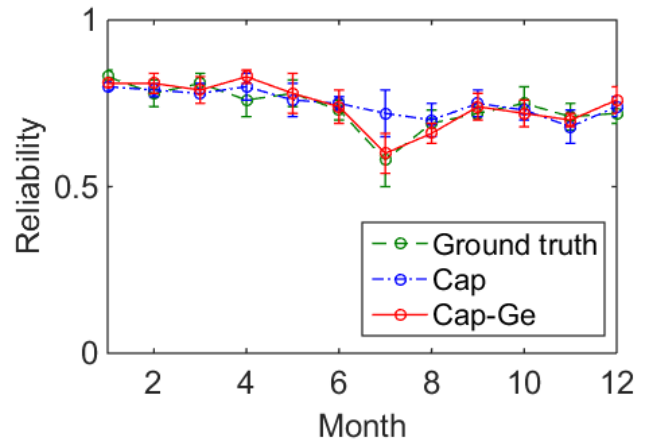


Fig. 5. Reliability evaluation by different methods over 12 months. We compare mean reliability as well as variation within one month here.

Note that, generated samples from decoder further prompt evaluation performance especially when drastic disturbances occur at equipments. We illustrate this difference in Fig. 5. One can see that without serious turbulence, two methods show approximately the same performance. However, due to some outside influence at July, there is an obvious reduction of reliability occurring. The model learned with proposed three stages strategy with generated samples follows this disturbance effectively.

V. CONCLUSION

In this paper, we address the reliability evaluation problem of power communication network equipments by introducing capsule networks from the field of NLP. To learn model with limited samples with heterogeneous structures, we proposed a three stages based learning framework. The framework exploit both original samples as well as generated samples by learned decoder with perturbation. Our

experimental results demonstrate that, capsule network is capable to learn deep models with small amount of samples and delivers acceptable performance in regression tasks such as reliability evaluation. We believe the proposed method also has extensive potential applications in various practical machine learning scenarios due to its low requirement of training samples.

ACKNOWLEDGMENT

This work was supported by Science and Technology Project of State Grid Jiangsu Electric Power Co., Ltd. in 2018. The project name is "Research on Fault Prediction and Network Quality Management and Control Technology of Power Communication Network Equipment Based on State Perception" (No. J2018054).

REFERENCES

- [1] L. Andersson, K.-P. Brand, C. Brunner, and W. Wimmer, "Reliability investigations for sa communication architectures based on iec 61850," in 2005 IEEE Russia Power Tech. IEEE, 2005, pp. 1–7.
- [2] Z.-D. Zhao, Y.-Y. Lou, J.-H. Ni, and J. Zhang, "Rbf-svm and its application on reliability evaluation of electric power system communication network," in 2009 International Conference on Machine Learning and Cybernetics, vol. 2. IEEE, 2009, pp. 1188–1193.
- [3] A. A. Leema and M. Hemalatha, "Proposed prediction algorithms based on hybrid approach to deal with anomalies of rfid data in healthcare," Egyptian Informatics Journal, vol. 14, no. 2, pp. 135–145, 2013.
- [4] Z. Xuegong, "Introduction to statistical learning theory and support vector machines," Acta Automatica Sinica, vol. 26, no. 1, pp. 32–42, 2000.
- [5] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," IEEE transactions on systems, man, and cybernetics, vol. 21, no. 3, pp. 660–674, 1991.
- [6] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in Advances in Neural Information Processing Systems, 2017, pp. 3859–3869.
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in neural information processing systems, 2017, pp. 5998–6008.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [9] W. Zhao, H. Peng, S. Eger, E. Cambria, and M. Yang, "Towards scalable and reliable capsule networks for challenging nlp applications," arXiv preprint arXiv: 1906.02829, 2019.
- [10] V. Jayasundara, S. Jayasekara, H. Jayasekara, J. Rajasegaran, S. Seneviratne, and R. Rodrigo, "Textcaps: Handwritten character recognition with very small datasets," in 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019, pp. 254–262.
- [11] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming autoencoders," in International Conference on Artificial Neural Networks. Springer, 2011, pp. 44–51.
- [12] S. Zhang, W. Zhao, X. Wu, and Q. Zhou, "Fast dynamic routing based on weighted kernel density estimation," arXiv preprint arXiv:1805.10807, 2018.
- [13] E. L. Mencia and J. F. Urnkranz, "Efficient pairwise multilabel classification for large-scale problems in the legal domain," in Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, 2008, pp. 50–65.
- [14] J. Liu, W.-C. Chang, Y. Wu, and Y. Yang, "Deep learning for extreme multi-label text classification," in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2017, pp. 115–124.