focused on predefined category classification. We here compare our model with three representative methods:
Literal Matching, where the product category is recognized from the product content by matching the exact words. If a word in the content matches an item in the given product category dictionary, then it is designated as a category of the product.
BiLSTM-CRF is favored in many content understanding tasks [12].
DeepCN [5] is our implementation of the state-of-the-art in large scale item categorization by product classification. The results are presented in Table 3.
When comparing the failure rate, we find that, when only recognizing the product categories from contents (as in Literal Matching, BiLSTM-CRF, and NPC-*Reco*), the models fails to complete the product categorization occasionally. They tend to not recognize any word as product category in some cases. Whereas, product categorizing by classification (DeepCN, NPC-*Pred*,NPC) would not suffer from similar problem.
In terms of the number of generated categories, Literal Matching and BiLSTM-CRF output approximately only one category for each product, while product category prediction models (DeepCN and NPC-*Pred*) generate more categories. The number of categories generated by NPC is larger than product category recognition models and fewer than product category prediction models. This illustrates that product category prediction tends to produce more fine-grained categories than product category recognition.
With respect to the evaluation metrics, we list the top one performance of the product category prediction models to perform a fair comparison. BiLSTM-CRF model performs better than Literal Matching. Literal Matching is prone to the fake categories. The precision of DeepCN is slightly lower than BiLSTM-CRF, however, the recall of DeepCN is much higher. All three variants of our model perform much better than all the baselines, while NPC-*Reco* achieves the best performance. This shows that our model is more adept at recognizing the product categories from product content.
Performance on Different Buckets. We further study the performance of the model with respect to different product exposure rates. From Figure 2, we notice that, our model consistently performs better than the baselines on all buckets. Products in bucket 0 are of relatively higher exposure rates than all other buckets. Both DeepCN and NPC surpasses BiLSTM-CRF by a large margin in bucket 0. As the exposure rates decreases, the performance of DeepCN gradually declines and BiLSTM-CRF overtakes DeepCN. The products in bucket 9 are long-tail products. NPC still performs much better than DeepCN and BiLSTM-CRF. We speculate that for high exposure rates product, product category prediction is a better choice, while for long-tail products, recognizing category from the product content is a superior option.

5 RELATED WORK

Product categorization problem in E-commerce is usually defined as classification of products into an existing list tens to thousands of categories [2, 3, 8]. Since the training data, usually consist of product titles and descriptions, are too sparse to be directly adopted as features, standard classification approaches utilized various types of manual features. [13] employed SVM in item categorization with unigram words, morphological features, quantity features and pattern features. A linear model is used to learn the multi-class classification problem with LDA and embedding features for Yahoo! products in [11]. [4] developed a two-level ensemble approach utilizing path-wise, node-wise and depth-wise classifiers based on document embedded vectors. [1] ensembled deep belief nets and KNN learned from semantic hashed product title and description to classify products. [15] illustrated that an attentional CNN with both word and character encodings increases the training efficiency, while [5] utilized a RNN. These approaches are only suitable for well-defined product classification with a limited number of categories. In fine-grained product categorization, each product belongs to multiple equivalent categories. We generate the product categories by jointly recognizing and predicting the product categories.

6 CONCLUSION

In this paper, we propose a neural product categorization model to generate the fine-grained categories from product contents by jointly recognizing from the product content and predicting from a predefined product category vocabulary. Moreover, to alleviate human labors, the proposed model is able to effectively adapt weak labels which is generated by mining search logs. Experiments on large-scale datasets from a real e-commerce platform demonstrate the effectiveness of the model. We'd like to perform more detailed comparisons and analysis in the future work.

REFERENCES

[1] Ali Cevahir and Koji Murakami. 2016. Large-scale Multi-class and Hierarchical Product Categorization for an E-commerce Giant. In *COLING*. 525–535.
[2] Eli Cortez, Mauro Rojas Herrera, Altigran S da Silva, Edleno S de Moura, and Marden Neubert. 2011. Lightweight methods for large-scale product categorization. *Journal of the American Society for Information Science and Technology* 62, 9 (2011), 1839–1848.
[3] Ying Ding, M Korotkiy, Borys Omelayenko, V Kartseva, V Zykov, Michel Klein, Ellen Schulten, and Dieter Fensel. 2002. Goldenbullet: Automated classification of product data in e-commerce. In *BIS*.
[4] Vivek Gupta, Harish Karnick, Ashendra Bansal, and Pradhuman Jhala. 2016. Product classification in e-commerce using distributional semantics. *arXiv* (2016).
[5] Jung-Woo Ha, Hyuna Pyo, and Jeonghee Kim. 2016. Large-scale item categorization in e-commerce using multiple recurrent neural networks. In *SIGKDD*. ACM, 107–115.
[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *CVPR*. 770–778.
[7] Dustin Hillard, Stefan Schroedl, Eren Manavoglu, Hema Raghavan, and Chirs Leggetter. 2010. Improving ad relevance in sponsored search. In *WSDM*. ACM, 361–370.
[8] Sheng Huang, Xinlan Liu, Xueping Peng, and Zhendong Niu. 2012. Fine-grained product features extraction and categorization in reviews opinion mining. In *ICDM Workshops*. IEEE, 680–686.
[9] Bhargav Kanagal, Amr Ahmed, Sandeep Pandey, Vanja Josifovski, Jeff Yuan, and Lluis Garcia-Pueyo. 2012. Supercharging recommender systems using taxonomies for learning user purchase behavior. *VLDB* 5, 10 (2012), 956–967.
[10] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv* (2014).
[11] Zornitsa Kozareva. 2015. Everyone likes shopping! multi-class product categorization for e-commerce. In *NAACL*. 1329–1333.
[12] Xuezhe Ma and Eduard Hovy. 2016. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. In *ACL*. Association for Computational Linguistics, Berlin, Germany, 1064–1074. http://www.aclweb.org/anthology/P16-1101
[13] Dan Shen, Jean David Ruvini, Manas Somaiya, and Neel Sundaresan. 2011. Item categorization in the e-commerce domain. In *CIKM*. ACM, 1921–1924.
[14] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. 2015. Highway networks. *arXiv* (2015).
[15] Yandi Xia, Aaron Levine, Pradipto Das, Giuseppe Di Fabbrizio, Keiji Shinzato, and Ankur Datta. 2017. Large-Scale Categorization of Japanese Product Titles Using Neural Attention Models. In *EACL*, Vol. 2. 663–668.