論文の内容の要旨

論文題目　Using Deep Learning to Make Data Manifest Knowledge
（データに知識を説明させるための深層学習活用）

氏　　名　張　确軒

1. Introduction

In the Industry 4.0 [1] and Society 5.0 [2] concepts, data is still the core component for the highly expected technologies, such as cyber-physical systems, the internet of things, artificial intelligence, and big data analytics. Deep learning (DL) [3] has been widely used as a versatile and high-performance tool to extract useful knowledge and information from data for problem-solving in various domains. By participating in the IMDJ [4] workshops aiming at better data utilization, we found that people criticize DL's poor interpretability. However, some data is difficult to extract features or label targets with human work, such as Fourier speckle patterns captured from the laser machining process [5]. DL is still required for those complicated problems. It motivated us to give a framework that facilitates the knowledge discovery process by using explanatory visualization on deep models with high accuracy and generalization. Here, knowledge is defined as useful information for problem-solving. For the sake of serving this framework, two model interpretation methods for deep models and a multi-task self-supervised approach for anomaly detection were proposed as follows. Also, a case study on the mentioned laser machining data was conducted.

2. Model Interpretation for Deep Models

Model interpretation (MI) is to explain the prediction processes of machine learning models for humans. Due to the "black-box" trait of deep models, it is essential to help people understand what deep models have learned or consider their reliability and improvement by using MI. We can typically group MI approaches into two categories, model-agnostic and model-specific [6]. Model-agnostic methods test the relevancy between input and output in a forward style without the model's architecture information, while model-specific methods must use the model architecture to propagate the relevancy from output to input backward. To enhance the interpretability of deep models, we proposed two MI methods, nonlinearized relevance propagation (NRP) [7] and key input subset sampling (KISS) [8].

NRP is altered from the previous method the layer-wise relevance propagation (LRP) [9]. LRP is a Taylor decomposition-based approach in which the output is resolved into relevance scores, and then the scores are backward propagated to the inputs layer-by-layer conservatively. In LRP, $\alpha\beta$-rules are used to control the ratio of positive and negative relevance in the decomposition for neural network layers. However, with the development of deep learning methods, many newly designed special layers require new decomposition rules in the use of LRP. For the pooling-weight layers, NRP introduced nonlinear functions into the relevance decomposition. The reasons why we considered nonlinear functions were: 1) those rules which do not satisfy the Taylor assumption could produce relevance errors in the propagation process, 2) the errors could accumulate layer-by-layer to the neurons at shallower layers, and 3) an appropriate nonlinear function could amend the relevance distribution to a more meaningful value space. Also, we provided a conservational $\alpha\beta$-rule for Hadamard product layers. Then, we applied NRP to interpret a question answering model, Attentive Pooling Network (APN) [10]. In the evaluation, we employed the token deleting test to compare our method with LRP. We found that applying certain nonlinear functions helped the explanation capture more important inputs than the linear setting.

KISS is based on the energy-based model (EBM) theory [11]. It can be a model-agnostic forward approach or a hybrid method by collaborating with backward information. Based on the EBM theory, we assume that each input element does positive or negative work for the prediction and estimate the work with the negative energy and the free energy. Considering the free energy also make the method overlooks all the relations between the input and each candidate answer. The EBM-based importance score (EBIS) is defined as the expectation of the work for each input element. In estimating EBIS, we use the importance sampling technique for the input subset sampling and employ gradient information as the sampling weights. In the evaluation, KISS outperformed other contrastive models on the image classification models. Also, we conducted a questionnaire survey to investigate the effect of MI's helping knowledge discovery on images by using the visualizational interpretation results by KISS on six images. According to the 31 responses, we found that the interpretation helped people notice blind spots, reasons of the intuition, or defects of the models. Therefore, we can expect MI approaches to facilitate the reflection upon the deep models and novelty discovery.

3. Case Study: Deep Learning on Laser Machining

In the case study [12], since the utilization of Fourier speckle patterns acquired from the laser machining process is still less studied in physics and data science, we firstly apply principal component analysis (PCA) on the dataset for the exploratory analysis. Then, to evaluate the performance of feature extraction on the speckle pattern data, we designed two tasks, Power Setting Classification (PSC) and Shot Number Regression (SNR). We utilized the cross-entropy loss [3] as the objective for the PSC, while applied smooth L_1 loss [13] as the one for SNR. The multi-task object is the sum of the two losses. We adopted AlexNet [14] and ResNet [15] for different models with the single-task or multi-task objective for the deep feature extraction. In the evaluation, we also used support vector machine (SVM) and simple fully connected neural networks as the baselines. According to the results, we found that using AlexNet for the feature extraction in MTL was better than the other comparative models for both tasks. However, the supervised DL approaches of feature extraction could weaken generalization for downstream tasks, since it is hard to label the anomalies from the sequential laser machining data by the recognized information.

4. Sequential Anomaly Detection with Pessimistic Contrastive Learning

To solve the difficulty in labeling anomalies from high-dimensional sequential data, I proposed pessimistic contrastive learning (PCL). In PCL, we pessimistically assume that there are anomalous data points in the sequences. The anomalies will then be recognized by driving the data points to contrast with each other within windows of context. The contrastive approach in PCL was inspired by the simple framework for contrastive learning of visual representations (SimCLR) [16], which maximizes the similarity of the two augmented data pairs' projections from the identical sample with the NT-Xent loss. In the sequential data modeling, I use these three assumptions: 1) the event that an anomaly occurs on a data point is independent of the other anomaly occurrences in the sequence, 2) the event that two of the data points have relation is independent of the other relationships in the sequence, and 3) a data point loses the relationship with the others if any anomaly occurs on it. Also, I designed the context-attempered relative entropy loss (CARE) to learn the anomaly-considered relations, the sequential NT-Xent loss (SeNT-Xent) to reconcile the augmented data in sequence, and the self-supervised sequential anomaly detection network (S^{3}ADNet) to predict the probability of anomaly occurring by using the multi-conceptual context (MCC) layer to capture the contextual information. In the experiments, I compared PCL to two commonly used anomaly detection methods on one-dimensional synthetic data and then applied PCL to find illegible handwritten digits from the MNIST dataset. The experiments proved that PCL could give meaningful results.

5. Limitations and Future Works

Nevertheless, there are still problems and limitations in the proposed methods, such as the parameter selection in NRP and KISS, the non-convolutional network for the Fourier data, and the hyperparameter optimization in PCL. Besides solving those problems in the future, we need to do more systemic work to facilitate humans' collaboration with deep learning for knowledge discovery.

References

[1] Klaus Schwab. The fourth industrial revolution. Currency, 2017.

[2] Mayumi Fukuyama. Society 5.0: Aiming for a new human-centered society. Japan Spotlight, 27:47–50, 2018.

[3] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. Deep learning, volume 1. MIT press Cambridge, 2016.

[4] Yukio Ohsawa, Hiroyuki Kido, Teruaki Hayashi, Chang Liu, and Kazuhiro Komoda. Innovators marketplace on data jackets, for valuating, sharing, and synthesizing data. In Knowledge-Based Information Systems in Practice, pages 83–97. Springer, 2015.

[5] Shuntaro Tani, Yutsuki Aoyagi, and Yohei Kobayashi. Neural-network-assisted in situ processing monitoring by speckle pattern observation. Optics Express, 28(18):26180–26188, 2020.

[6] Christoph Molnar. Interpretable Machine Learning. Lulu. com, 2020.

[7] Quexuan Zhang and Yukio Ohsawa. Nonlinearized relevance propagation. In Pacific Rim International Conference on Artificial Intelligence, pages 904–914. Springer, 2018.

[8] Quexuan Zhang and Yukio Ohsawa. Kiss: an ebm-based approach for explaining deep models. Procedia Computer Science, 176:271–280, 2020.

[9] Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one, 10(7):e0130140, 2015.

[10] Ming Tan, Cicero Dos Santos, Bing Xiang, and Bowen Zhou. Improved representation learning for question answer matching. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 464–473, 2016.

[11] Yann LeCun, Sumit Chopra, Raia Hadsell, M Ranzato, and F Huang. A tutorial on energy-based learning. Predicting structured data, 1(0), 2006.

[12] Quexuan Zhang, Zexuan Wang, Bin Wang, Yukio Ohsawa, and Teruaki Hayashi. Feature extraction of laser machining data by using deep multi-task learning. Information, 11(8):378, 2020.

[13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 91–99, 2015.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6):8490, 2017.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[16] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. arXiv preprint arXiv:2002.05709, 2020.