Maja Rudolph | Research Statement

Increases in computing power have led to impressive leaps in the performance of deep learning systems. However, these methods are not yet reliable enough to be deployed in many critical applications. Even when trained on enormous amounts of clean and labeled data, they fail in unpredictable ways when the deployment environment changes. Moreover, they often assume that the training data is i.i.d. (independently and identically distributed), while in many important settings the data is correlated in interesting ways.

This raises the following questions: How can we learn from unlabeled and potentially contaminated data? How can we distill the information encoded in a dataset into reusable representations? How can we combine the flexibility of neural networks with principles of probabilistic modeling to learn distributions over structured data such as time series?

To answer these questions, I develop novel machine-learning methods at the intersection of deep learning and probabilistic modeling. There are two main threads in my research:

- In my first research direction, **deep probabilistic sequence models**, I study how to best combine probabilistic modeling and deep learning to model time-series data.
- In my second research direction, **neural transformations and embeddings**, I develop unsupervised and self-supervised methods for anomaly detection and representation learning. These methods are particularly suited for unlabeled data.

Research Direction 1: Deep Probabilistic Sequence Models

Many types of data reveal themselves to us sequentially. Examples include video, neural firing patterns, financial market data, or sequential sensor measurements in engineering. Machine-learning methods for such data can lead to new scientific insights, aid decision making, or increase the utility of engineering devices by improving their efficiency or enabling new features. However, learning from sequential data requires methods that are flexible enough to capture complex relationships in the data (deep learning) while handling various types of noise and uncertainty (probabilistic modeling).

Below are two specific examples of deep probabilistic sequence models that I developed jointly with my students to address challenges arising from industrial applications.

Modeling irregular time series. Common issues in dynamics modeling for industrial applications are that different sensors are usually not synchronized, sampling rates can be irregular, and some sensors are only partially observed. A typical approach is to interpolate the inputs before applying a standard deep-learning method, such as a recurrent neural network (RNN). The *continuous recurrent unit* (CRU) **[ICML 2022a]** does not require this heuristic pre-processing step. Like an RNN, it maintains a latent state which is recursively updated based on observations. After each observation is encoded by a neural network, the CRU uses the continuous formulation of a Kalman filter to optimally integrate the new information into its latent state. The result is a practical neural architecture that handles irregularly sampled time-series data naturally.

Forecasting multimodal dynamics. Another modeling challenge is that many real-world time-series datasets are highly multimodal: at any given part of an observed sequence, there are multiple plausible continuations (e.g. a car could turn right or left at an intersection) but the average of these behaviors is unlikely or even physically impossible. Most deep probabilistic sequence models struggle with multimodality since their modeling assumptions and training objectives encourage averaging. To address this, we have developed a new approximate inference method for sequential latent variables that is multimodal by construction **[Entropy 2021]**. The key idea is to marginalize over past states instead of conditioning on them, as in previous work. Our approach improves forecasting in several real-world datasets, including taxi trajectories, movement patterns of basketball players, and pollution levels over time.

Research Direction 2: Neural Transformations and Embeddings

Pressing challenges for machine-learning research can also be encountered in manufacturing. Production facilities generate large amounts of sensor data, which can be processed with machine learning to make production more efficient or to automate quality inspection.

Data collected during day-to-day usage of an assembly line will typically not be labeled. For this reason, unsupervised and self-supervised methods are needed to make sense of production data. I will describe two classes of methods from this category: self-supervised anomaly detection and embeddings for learning interpretable representations.

2.1. Neural Transformations for Anomaly Detection

In self-supervised learning, auxiliary tasks replace labels as the prediction targets. Models trained on these tasks can then be used for anomaly detection or feature extraction. While there has been extensive work in self-supervised anomaly detection on images, many industrial applications require us to detect abnormalities in other types of data, such as time series (e.g. abnormal process curves of a drill), graphs (e.g. for detecting abnormal molecules in a chemical process), or events within a time series (e.g. for intrusion detection in a water distribution system).

Self-supervised learning in computer vision often relies on image transformations such as rotations, cropping, and blurring to construct auxiliary tasks. Applications of these methods to industrial data have been limited since it is difficult to manually design appropriate data transformations for specialized domains. To overcome this challenge, we have developed *neural transformations* **[ICML 2021, ICML 2022b]**. We treat the transformations like model parameters and thereby can *learn* appropriate transformations from the data.

Our main contribution is a novel training objective that encourages the neural transformations to generate alternative views of the data that are diverse, yet characteristic of the original input. We prove that optimizing this objective avoids the trivial solutions (identity or constant) that one would encounter with existing losses. Neural transformations achieve state-of-the-art results for anomaly detection of entire time series and tabular data **[ICML 2021]**, of windows within time series **[SSL 2021]**, and on graphs **[IJCAI 2022]**.

Deep one-class classification (OCC) is another popular self-supervised approach for anomaly detection. While deep OCC works well on images, the optimal solution of the OCC objective is trivial and cannot be used for anomaly detection. We prove in **[IJCAI 2022]** that combining OCC with transformation learning completely resolves this issue. The theoretical findings are supported by our empirical results where our combined method, *one-class graph transformation learning*, achieves the highest accuracy in detecting abnormal graphs.

Neural transformations have another advantage. Since the data transformations are not crafted by hand but are instead learned by neural networks, we can apply data augmentations to data types that we have less intuition about than natural images. In fact, we can even apply the neural transformations in an abstract feature space [**NeurIPS 2017**, **SSL 2021**], sometimes called the *embedding space*. The embedding vector of each data point captures characteristics of the data that are useful for neural networks but not necessarily accessible to human interpretation. Their abstract nature makes it hard to design transformations by hand but lends itself to neural transformations.

2.2. Embedding Methods for Learning Interpretable Representations.

Embeddings are a powerful tool for working with high-dimensional data. Word embeddings, for example, capture semantic similarity among terms in a vocabulary and are useful inputs for downstream tasks **[PNAS 2020]**. To make them applicable to other types of data, we have developed *exponential family embeddings* **[NeurIPS 2016]**, which can be combined with a dynamic prior to study how similarities change over time **[WWW 2018]**. There is a great potential for employing embeddings in data-driven manufacturing. Embeddings that capture the characteristics of parts, stations, or production steps can simplify tasks such as root-cause analysis and reduce the amount of training data required for prediction.

Future Work

My work on time-series models has focused on integrating probabilistic approaches into recurrent neural networks to address challenges ranging from multi-modality **[Entropy 2021]** to irregularly sampled observations **[ICML 2022a]**. Other interesting challenges include the multi-scale setting, where the underlying dynamics have multiple time scales, and the continual learning setting, where the distribution associated with the deployment environment keeps shifting. In terms of method development, the probabilistic tools I have used to improve on RNNs could also be beneficial for other deep sequence models, such as transformers, which have been outperforming RNNs in many settings.

My work on neural transformations [**ICML 2021**, **ICML 2022b**, **IJCAI 2022**] enables self-supervised anomaly detection in specialized domains with data types beyond images. Another very important area is self-supervised representation learning. Here, the neural transformations we have developed are not yet on par with other techniques for data augmentation. Contrastive losses derived from alternative information-theoretic criteria may yield the right loss functions for learning neural transformations for representation learning. More powerful neural transformations would in turn lead to improved representation learning algorithms for less-studied data types including time series, tabular data, and graphs, which often occur in engineering, scientific, and medical applications.

The power of deep learning comes from its ability to distill information into distributed representations. These representations provide an embedding space in which many machine-learning tasks (e.g. prediction) are easier to solve. While there are analogies between embedding methods and models of language processing in the brain, some aspects of human cognition (common sense, hierarchical categorization, causal understanding, physical intuition) are not well reflected in current machine-learning embedding models **[PNAS 2020]**. I hope to address this in future work.

Selected Bibliography

[ICML 2022a] M. Schirmer, M. Eltayeb, S. Lessmann, and **M. Rudolph**. Modeling irregular time series with Continuous Recurrent Units. In *ICML*, 2022.

[ICML 2022b] C. Qiu, A. Li, M. Kloft, **M. Rudolph**, and S. Mandt. Latent outlier exposure for anomaly detection with contaminated data. In *ICML*, 2022.

[IJCAI 2022] C. Qiu, M. Kloft, S. Mandt, and **M. Rudolph**. Raising the bar in graph-level anomaly detection. In *IJCAI*, 2022.

[ICML 2021] C. Qiu, T. Pfrommer, M. Kloft, S. Mandt, and **M. Rudolph**. Neural transformation learning for deep anomaly detection beyond images. In *ICML*, 2021.

[Entropy 2021] C. Qiu, S. Mandt, and **M. Rudolph**. History marginalization improves forecasting in variational RNNs. In *Entropy*, 2021.

[SSL 2021] T. Schneider, C. Qiu, M. Kloft, D. Latif, S. Staab, S. Mandt, and **M. Rudolph**. Detecting anomalies in time series with local neural transformations. In *ICML Workshop: Self-Supervised Learning for Reasoning and Perception*, 2021.

[PNAS 2020] J. McClelland, F. Hill, **M. Rudolph**, J. Baldridge, and H. Schütze. Placing language in an integrated understanding system: Towards human-level performance in neural language models. In *Proceedings of the National Academy of Sciences*, 2020.

[WWW 2018] **M. Rudolph** and D. Blei. Dynamic embeddings for language evolution. In *World Wide Web Conference*, 2018.

[NeurIPS 2017] **M. Rudolph**, F. Ruiz, S. Athey, and D. Blei. Structured embedding models for grouped data. In *Neural Information Processing Systems*, 2017.

[NeurIPS 2016] **M. Rudolph**, F. Ruiz, S. Mandt, and D. Blei. Exponential family embeddings. In *Neural Information Processing Systems*, 2016.